

DEX-0257

CD

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/038606 A3(51) International Patent Classification⁷: C07K 14/47,
C12N 15/12, 15/10, C12Q 1/68, C07K 16/18, G01N
33/53, A61K 39/00(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(21) International Application Number: PCT/US01/46459

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.(22) International Filing Date:
7 November 2001 (07.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/246,640 8 November 2000 (08.11.2000) US(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/246,640 (CIP)
Filed on 8 November 2000 (08.11.2000)**Published:**

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(71) Applicant (*for all designated States except US*): DI-
ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): SUN, Yongming
[CN/US]; 551 Shoal Drive, Redwood City, CA 94065
(US). RECIPON, Herve [FR/US]; 85 Fortuna Avenue,
San Francisco, CA 94115 (US). SALCEDA, Susana
[AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136
(US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14,
San Jose, CA 95117 (US).(88) Date of publication of the international search report:
15 May 2003For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/038606 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/JP 01/46459

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/47 C12N15/12 C12N5/10 C12Q1/68 C07K16/18
G01N33/53 A61K39/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, SEQUENCE SEARCH, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 23 September 1998 (1998-09-23) ADAMS, MD ET AL.: "RPCI11-43J8.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-43J8, genomic" retrieved from EBI Database accession no. AQ194904 XP002232022 abstract	1-5,7,8
X	WO 00 11014 A (ENDRESS GREGORY A ;FLORENCE KIMBERLY A (US); HUMAN GENOME SCIENCES) 2 March 2000 (2000-03-02) SEQ ID NO:49, 170 page 127 -page 128 --- -/--	1-5,7-10

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

20 February 2003

Date of mailing of the international search report

13. 03. 2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Schmitz, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/... 01/46459

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 29 June 2000 (2000-06-29) PHILLIMORE B.: "Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-37015" retrieved from EBI Database accession no. AL360009 XP002217521 abstract	1-5
P,X	--- DATABASE EMBL [Online] 31 July 2001 (2001-07-31) KAUL RK ET AL.: "Homo sapiens chromosome 1 clone RP11-365016, complete sequence." retrieved from EBI Database accession no. AC092809 XP002232021 abstract	1-5,7,8
A	--- DATABASE SWALL [Online] 1 August 1998 (1998-08-01) MYLER P. ET AL.: "CAKC2" retrieved from EBI Database accession no. 060981 XP002217522 abstract	
A	--- WO 00 24755 A (BAYLOR COLLEGE MEDICINE ;MATZUK MARTIN M (US); WANG PEI (US)) 4 May 2000 (2000-05-04) -----	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/46459

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
1-17 (all partially with respect to inventions 1 and 33)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-17 (all partially)

Biosequence as defined by SEQ ID NO:1.
Furthermore, as far as applicable, vectors, host cells, methods, antibodies, kits, vaccines, all referring to said biosequence.

Since no clear relationship between the amino acid- and nucleic acid sequences could be established, all SEQ ID NOs are regarded as individual inventions.

Invention 2: claims 1-17 (all partially)

As invention 1, but referring to SEQ ID NO:2.

.
.
.

Invention 238: claims 1-17 (all partially)

As invention 1, but referring to SEQ ID NO:238.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 14 is directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

INTERNATIONAL SEARCH REPORT

nation on patent family members

International Application No

PCT/US 01/46459

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0011014	A	02-03-2000	AU 5583799 A	14-03-2000
			CA 2340884 A1	02-03-2000
			EP 1109821 A1	27-06-2001
			JP 2002523035 T	30-07-2002
			WO 0011014 A1	02-03-2000
			US 2001021700 A1	13-09-2001
			US 2002026040 A1	28-02-2002

WO 0024755	A	04-05-2000	AU 1326200 A	15-05-2000
			EP 1124840 A1	22-08-2001
			WO 0024755 A1	04-05-2000
			US 2002042926 A1	11-04-2002

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38606 A2

(51) International Patent Classification⁷: **C07K 14/47**

San Jose, CA 95136 (US). **LIU, Chenghua** [CN/US];
1125 Ranchero Way #14, San Jose, CA 95117 (US).

(21) International Application Number: **PCT/US01/46459**

(74) Agents: **LICATA, Jane, Massey et al.**; Licata & Tyrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date:
7 November 2001 (07.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/246,640 8 November 2000 (08.11.2000) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:

US 60/246,640 (CIP)
Filed on 8 November 2000 (08.11.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

(71) Applicant (*for all designated States except US*): **DI-
ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

Published:

— *without international search report and to be republished
upon receipt of that report*

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SUN, Yongming**
[US/US]; 869 S. Winchester Boulevard, Apartment 260,
San Jose, CA 95128 (US). **RECIPON, Herve** [FR/US];
85 Fortuna Avenue, San Francisco, CA 94115 (US).
SALCEDA, Susana [AR/US]; 4118 Crescendo Avenue,

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

WO 02/38606 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/246,640 filed November 8, 2000, which is herein incorporated by reference
in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic ovary cells, including fragments, variants
and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
15 of the invention and methods for the use of these compositions. These uses include
identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and
non-cancerous disease states in ovary tissue, identifying ovary tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
20 production of engineered ovary tissue for treatment and research.

BACKGROUND OF THE INVENTION

Cancer of the ovaries is the fourth-most cause of cancer death in women in
the United States, with more than 23,000 new cases and roughly 14,000 deaths
predicted for the year 2001. Shridhar, V. et al., Cancer Res. 61(15):
25 5895-904 (2001); Memarzadeh, S. & Berek, J. S., J. Reprod. Med. 46(7):
621-29 (2001). The incidence of ovarian cancer is of serious concern
worldwide, with an estimated 191,000 new cases predicted annually.
Runnebaum, I. B. & Stickeler, E., J. Cancer Res. Clin. Oncol. 127(2): 73-79
(2001). Because women with ovarian cancer are typically asymptomatic until
30 the disease has metastasized, and because effective screening for ovarian
cancer is not available, roughly 70% of women present with an advanced stage

-2-

of the cancer, with a five-year survival rate of ~25-30% at that stage.

Memarzadeh, S. & Berek, J. S., *supra*; Nunns, D. et al., *Obstet. Gynecol.*

Surv. 55(12): 746-51. Conversely, women diagnosed with early stage ovarian cancer enjoy considerably higher survival rates. Werness, B. A. &

- 5 Eltabbakh, G. H., *Int'l. J. Gynecol. Pathol.* 20(1): 48-63 (2001).

Although our understanding of the etiology of ovarian cancer is incomplete, the results of extensive research in this area point to a combination of

age, genetics, reproductive, and dietary/environmental factors. Age is a key risk factor in the development of ovarian cancer: while the risk for

- 10 developing ovarian cancer before the age of 30 is slim, the incidence of ovarian cancer rises linearly between ages 30 to 50, increasing at a slower rate thereafter, with the highest incidence being among septagenarian women.

Jeanne M. Schilder et al., *Hereditary Ovarian Cancer: Clinical Syndromes and Management*, in *Ovarian Cancer* 182 (Stephen C. Rubin & Gregory P. Sutton
15 eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the most significant risk factor in the development of the disease, with that risk depending on the number of affected family members, the degree of their relationship to the woman, and which particular first degree relatives are

- 20 affected by the disease. *Id.* Mutations in several genes have been associated with ovarian cancer, including BRCA1 and BRCA2, both of which play a key role in the development of breast cancer, as well as hMSH2 and hMLH1, both of which are associated with hereditary non-polyposis ovary cancer. Katherine Y. Look, *Epidemiology, Etiology, and Screening of Ovarian*
25 *Cancer*, in *Ovarian Cancer* 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on chromosome 13, are tumor suppressor genes implicated in DNA repair; mutations in these genes are linked to roughly 10% of ovarian cancers. *Id.* at 171-72; Schilder et al., *supra* at 185-86. hMSH2 and hMLH1 are associated with DNA
30 mismatch repair, and are located on chromosomes 2 and 3, respectively; it has been reported that roughly 3% of hereditary ovarian carcinomas are due to mutations in these genes. Look, *supra* at 173; Schilder et al., *supra* at

184, 188-89.

Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at menarche have all been linked with an elevated risk of ovarian cancer.

- 5 Schilder et al., *supra* at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. *Id.*; Laura J. Havrilesky & Andrew Berchuck, *Molecular Alterations in Sporadic Ovarian Cancer*, in *Ovarian*
10 *Cancer* 25 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undesired mutation will occur. *Id.* Support for this theory may be found in the fact pregnancy,
15 lactation, and the use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. *Id.*

- Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the
20 antioxidant Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. *Look, supra* at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. *Id.* at 169-70.

- 25 Current screening procedures for ovarian cancer, while of some utility, are quite limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, *Cancer: Etiology, Diagnosis, and Treatment* 166 (1998); Memarzadeh & Berek, *supra*;
30 Runnebaum & Stickeler, *supra*; Werness & Eltabbakh, *supra*. Commonly used screening tests include bimanual rectovaginal pelvic examination, radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal

ultrasonography. Burdette, *supra* at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. *Id.* One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed with the disease at the time of their pelvic examination. Look, *supra* at 174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. *Id.* at 174-75. Although transvaginal ultrasonography is now the preferred procedure for screening for ovarian cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., *supra* at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for some patients and may also yield false negative or indeterminate results. Schilder et al., *supra* at 191-94.

The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, *supra* at 170; Memarzadeh & Berek, *supra*; Shridhar et al., *supra*. Staging is performed by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, Primary Surgical Management of Early Epithelial Ovarian Carcinoma, in Ovarian Cancer 203 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001); Fleming et al. eds., *supra* at 188. Stage I ovarian cancer is characterized by tumor growth that is limited to the ovaries and is comprised of three substages. *Id.* In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present in ascites or peritoneal washings. *Id.* Substage IB is identical to A1, except that tumor growth is limited to both ovaries. *Id.* Substage IC refers to the presence of tumor growth limited to one or both ovaries, and

also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. Id.

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along with pelvic extension. Id. Substage IIA involves extension and/or implants on the uterus and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. Id. Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings. Id.

Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. Id. Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal metastasis outside the pelvis 2 cm or less in greatest dimension. Id. Substage IIIC is identical to IIIB, except that the metastasis is greater than 2 cm in greatest dimension and may include regional lymph node metastasis. Id. Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. Id.

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, *supra* at 206-208, 213. In view of these limitations, attention has turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers to help better assess the progression of the disease. Vartiainen, J. et al., *Int'l J. Cancer*, 95(5): 313-16 (2001); Shridhar et al. *supra*; Baekelandt, M. et al., *J. Clin. Oncol.* 18(22): 3775-81.

The treatment of ovarian cancer typically involves a multiprong attack, with

-6-

surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, Primary Surgical Management of Advanced Epithelial Ovarian Cancer, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., Expert Op. Pharmacother. 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most patients experience tumor recurrence within three years of treatment. Id. Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, supra.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop ovarian cancer, for diagnosing ovarian cancer, for monitoring the progression of the disease, for staging the ovarian cancer, for determining whether the ovarian cancer has metastasized, and for imaging the ovarian cancer. There is also a need for better treatment of ovarian cancer.

30

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and
5 non-cancerous disease states in ovaries; identify and monitor ovary tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovary tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that
10 are specific to ovary cells and/or ovary tissue. These ovary specific nucleic acids (OSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is an ovary specific gene (OSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to
15 ovary. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 138 through 238. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 137. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence
20 similarity to a nucleic acid molecule encoding an OSP, or that selectively hybridize or exhibit substantial sequence similarity to an OSNA, as well as allelic variants of a nucleic acid molecule encoding an OSP, and allelic variants of an OSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an OSP or that comprises a part of a nucleic acid sequence of an OSNA are also provided.

25 A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an OSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment
30 of an OSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic

acid molecule encodes all or a fragment of an OSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an OSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly
5 produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an OSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by
10 an allelic variant of an OSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

15 Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and
20 treating ovarian cancer and non-cancerous disease states in ovaries. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring ovary tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered ovary tissue for treatment and research.

25 The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries. The invention provides methods of using the polypeptides of the invention to identify and/or monitor ovary tissue, and to produce engineered ovary tissue.

30 The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in ovaries and to produce engineered ovary tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

5 DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular
10 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed
15 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor
20 Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of
25 which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and
30 techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 5 A “nucleic acid molecule” of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A “nucleic acid molecule” as used herein is synonymous with “nucleic acid” and
- 10 “polynucleotide.” The term “nucleic acid molecule” usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 15 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 20 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially
- 25 duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.
- 30 A “gene” is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term “exon” refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term “intron” refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be “spliced out” during processing of the transcript.

10 A nucleic acid molecule or polypeptide is “derived” from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An “isolated” or “substantially pure” nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the “isolated polynucleotide” is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term “isolated” or “substantially pure” also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term “isolated nucleic acid molecule” includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

30 A “part” of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, e.g. for use as probes or primers, or may be double-stranded, e.g. for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized
5 oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate,
15 phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No.
20 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is
25 referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of
30 RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

-16-

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

5 $T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35$
(% formamide) - (820/ l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50$$

(% formamide) - (820/ l).

10 In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C
15 would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.*

- 5 (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- 15 As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- 20 Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-30 11.57.

The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μ l of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an OSP or is an OSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term “sexual PCR mutagenesis” or “DNA shuffling” refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes (“Family shuffling”).

10 The term “*in vivo* mutagenesis” refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

20 The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

25 The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

5 The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,
10 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such
15 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20 The term “vector,” as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of
25 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and
30 thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply, “expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector.

However, the invention is intended to include other forms of expression vectors that
5 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding
10 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in
15 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-
25 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an OSP encoded by a nucleic acid molecule of the instant
30 invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be

5 “isolated” from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is “substantially pure,” “substantially homogeneous” or “substantially purified” when at least about 60% to 75% of a sample exhibits a single

10 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

15 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term “polypeptide fragment” as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared

20 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

25 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A “derivative” refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such

30 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

-27-

acids may also be suitable components for polypeptides of the present invention.

Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate,

-N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine,

N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other

5 similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism
10 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"
15 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence
20 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino
25 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino
30 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.,* Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are
5 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A “moderately conservative” replacement is
15 any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid
20 substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.,* GCG Version 6.1. Other programs include FASTA, discussed *supra*.

25 A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

30 Expectation value: 10 (default)
Filter: seg (default)
Cost to open a gap: 11 (default)

-29-

Cost to extend a gap: 1 (default)
 Max. alignments: 100 (default)
 Word size: 11 (default)
 No. of descriptions: 100 (default)
 5 Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number
 10 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best
 15 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

20 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,
 25 dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab
 30 fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

5 A “neutralizing antibody” or “an inhibitory antibody” is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An “activating antibody” is an antibody that increases the activity of a polypeptide.

10 The term “epitope” includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1 μ M, preferably less than 100 nM and most preferably less than 10 nM.

15 The term “patient” as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20 The term “ovary specific” refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovary as compared to other tissues in the body. In a preferred embodiment, a “ovary specific” nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the “ovary specific” nucleic acid molecule or polypeptide is
25 expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,
30 such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

5 One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovary or to ovary cells or tissue or that are derived from such nucleic acid molecules. These isolated ovary specific nucleic acids (OSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid
10 molecule encodes a polypeptide that is specific to ovary, an ovary-specific polypeptide (OSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 138 through 238. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 137.

15 AN OSNA may be derived from a human or from another animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

 By "nucleic acid molecule" for purposes of the present invention, it is also meant
20 to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an OSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that
25 selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 138 through 238. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 137.

30 In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 138 through 238. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 137. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an OSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human OSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 138 through 238. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 138 through 238, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an OSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an OSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 137. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 137, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an OSNA, more preferably at least 95%,
5 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an OSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one
10 that exhibits sequence identity over its entire length to an OSNA or to a nucleic acid molecule encoding an OSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding an OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at
15 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 138 through 238 or demonstrates
20 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 137. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated
25 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other
30 species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is

experimentally produced by directed mutation of an OSNA. Further, the substantially similar nucleic acid molecule may or may not be an OSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an OSNA or a nucleic acid encoding an OSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 137. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an OSP. However, in a preferred embodiment, the part encodes an OSP. In one aspect, the invention comprises a part of an OSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an OSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an OSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an OSP. A part comprises at least 10 nucleotides, more preferably at least 15,

17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that
5 encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain
10 reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains
15 modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe,
20 the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be
25 limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment,
30 the labeled nucleic acid molecule may be used as a hybridization probe.

-37-

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{-32}P -dATP, ^{-32}P -dCTP, ^{-32}P -dGTP, ^{-32}P -dTTP, ^{-32}P -3'dATP, ^{-32}P -ATP, ^{-32}P -CTP, ^{-32}P -GTP, ^{-32}P -UTP, ^{-35}S -dATP, α - ^{35}S -GTP, α - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al., J. NIH Res.* 5: 82 (1994); Van Belkum *et al., BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g., Tyagi et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al., Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al., Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al., Science* 279: 1228-1229 (1998); Marras *et al., Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al., Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al., Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al., Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science*, Kluwer Law International (1999); Stein *et al. (eds.), Applied Antisense Oligonucleotide Technology*, Wiley-Liss (1998); Chadwick *et al. (eds.),*

Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000),
5 the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including
10 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above
15 phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by
20 reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or
25 more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and
30 methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S.

Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases

do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.,* Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably,

detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.,* Andreeff *et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify OSNA in, and isolate OSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.,* Schwarczacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in
5 their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an OSP. In a more
10 preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 138 through 238. In another preferred embodiment, the probe or primer is derived from an OSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 137.

15 In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer
20 in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide
25 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic
30 Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular

Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention

can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides

encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones
5 *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000);
10 Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an
15 expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part
20 of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic
25 nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the
30 nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their

-47-

derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*,
5 typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic
10 genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will
15 typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2,
20 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*,
25 *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors
30 are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-

transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

5 In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines
10 expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian
15 cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

20 Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

25 Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

 It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for
30 encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of

the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these
5 vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include
10 splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that
15 modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC
20 system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

25 Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

30 Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early

gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the
5 promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include
10 introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may
15 alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are
20 described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors
25 include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the
30 PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to

be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

-52-

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.*, Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/ResearchCommittees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one
5 may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is
10 capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

15 In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the
20 product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression
25 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have
30 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell
5 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an
10 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be
15 able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*
20 *frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from
25 *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3
30 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from ovary are particularly preferred because they may provide a more native
5 post-translational processing. Particularly preferred are human ovary cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number
10 of texts and laboratory manuals in the art. *See, e.g.,* Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend
15 primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.,* Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

20 Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.,* with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent
25 strains are also available commercially (*e.g.,* Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse
30 treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the
5 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective
15 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (<http://www.bio-rad.com/LifeScience/pdf/>

- 5 New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 10 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid 15 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), 20 Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by 25 means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

- 30 Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is an ovary specific polypeptide (OSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 138 through 238. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an OSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 138 through 238. A polypeptide that comprises only a fragment of an entire OSP may or may not be a polypeptide that is also an OSP. For instance, a full-length polypeptide may be ovary-specific, while a fragment thereof may be found in other tissues as well as in ovary. A polypeptide that is not an OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. However, in a preferred embodiment, the part or fragment is an OSP. Methods of determining whether a polypeptide is an OSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

-64-

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies
5 (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

10 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75
15 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically
20 synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a
25 fragment of polypeptide of the invention, preferably an OSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an OSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants,
30 fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be ovary-specific. In a preferred embodiment, the mutein is ovary-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 138 through 238. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovary-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel 5 (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an OSP. In an even more preferred embodiment, the 10 polypeptide is homologous to an OSP selected from the group having an amino acid sequence of SEQ ID NO: 138 through 238. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an OSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 138 through 238. In an even 15 more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, 20 more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an OSP comprising an 25 amino acid sequence of SEQ ID NO: 138 through 238. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an OSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that 30 hybridizes to an OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSNA is selected from the group consisting of SEQ ID NO: 1 through 137. In another preferred

embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an OSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSP is selected from the group consisting of SEQ ID NO: 138 through
5 238.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of
10 SEQ ID NO: 138 through 238. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster,
15 cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous
20 polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an OSP. Further, the
25 homologous protein may or may not encode polypeptide that is an OSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an
30 antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the

-68-

binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an OSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 138 through 238. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 137.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 138 through 238, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, e.g., radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546; Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 5 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, 10 HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, 15 SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

20 The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

25 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

30 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 138 through 238. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- 15 --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000);
- 30 Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added
5 using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during
10 synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be
15 incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side
20 chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-
25 aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic
30 acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

-73-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 138 through 238, or is a mutein, homologous

polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 137, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule

5 comprising a nucleic acid sequence of SEQ ID NO: 1 through 137.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino

10 acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and

15 usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety,

20 heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further

25 purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion

30 proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See Bartel et al. (eds.), The Yeast Two-Hybrid System*, Oxford University Press (1997); Zhu *et al.*, *Yeast Hybrid Technologies*, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

-76-

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded
5 protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin
10 A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin *trpE*, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine
15 at the amino and/or carboxyl terminus of the polypeptide), *lacZ*, green fluorescent protein (GFP), yeast α -mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by
20 enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind
25 the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize OSPs, their allelic
30 variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser

scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

5 One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon
10 linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-
15 102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987).
20 Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-
25 proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For
30 example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is an OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 138 through 238, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on an OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of an OSP may be indicative of cancer. Differential degradation of the C or N-terminus of an OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human ovary.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

-80-

present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as
5 IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such
10 antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic
15 mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by
20 reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger
30 mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

-81-

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, 5 *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their 10 entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding 15 antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 20 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody 25 fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein 30 (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

- 5 Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab
10 fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- 15 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3): 157-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*,
20 *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.*
25 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

- Antibodies and fragments and derivatives thereof of the present invention can
30 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

Biotechnol. Appl. Biochem. 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entirety.

5 Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entirety.

10 Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression
15 systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57
20 (1999), the disclosures of which are incorporated herein by reference in their entirety.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the
25 proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

30 It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated

nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

5 Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

10 Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature*
15 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2
20 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entirety.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain
25 diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof,
30 whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

10 As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

20 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,
25 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,
30 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H ,
5 and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

10 As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the
15 application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria*
20 toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate,
25 and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more
30 of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to
5 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in
10 prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present
20 invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody
25 molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human
30 organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an OSP. In a preferred embodiment, the OSP comprises an amino

-90-

acid sequence selected from SEQ ID NO: 138 through 238, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an OSNA of the invention, preferably an OSNA comprising a nucleotide sequence selected from the group
5 consisting of SEQ ID NO: 1 through 137, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-
10 human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al., Biotechnology* 11:
20 1263-1270 (1993); Wright *et al., Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.*
25 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of
30 nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g., Campbell et al., Nature* 380: 64-66 (1996); Wilmut *et al., Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*

-91-

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene
5 may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

10 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using
15 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

20 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce ovaries of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels
25 because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is
30 appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

-92-

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt
5 the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

10 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

15 Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

20 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the
25 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such
30 conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 137 and SEQ ID NO: 138 through 238 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Ovarian Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of an OSNA or an OSP in a human patient that has or may have ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of an OSNA or an OSP in a normal human control. For purposes of the present invention, “expression of an OSNA” or “OSNA expression” means the quantity of OSG mRNA that
5 can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term “expression of an OSP” or “OSP expression” means the amount of OSP that can be measured by any method known in the art or the level of translation of an OSG OSNA that can be measured by any method known in the art.

10 The present invention provides methods for diagnosing ovarian cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an OSNA or
15 OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of an OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing,
20 alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in an OSP compared to an OSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the OSP or subcellular OSP localization.

25 In a preferred embodiment, the expression of an OSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 138 through 238, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of an OSNA mRNA selected from SEQ ID NO: 1 through 137, or a
30 hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. OSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.,* Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter
5 gene hooked up to the promoter of an OSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.,* aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovary nucleic acid, to detect a change in expression.

10 In another preferred embodiment, the expression of an OSP is measured by determining the level of an OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 138 through 238, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for
15 instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of OSNA or OSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of an OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be
20 determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP
25 structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An
30 antibody specific to an OSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP detected which is attached to the solid support can be correlated to the quantity of an OSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of OSNA or OSP includes, without limitation, ovary tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, ovary cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and breast. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. See Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in OSNAs or OSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

-100-

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP and then ascertaining whether the patient has ovarian cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a variety of tissues. The presence of an OSNA or OSP in a

-101-

certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of an OSNA or OSP is associated with ovarian cancer. Similarly, the presence of an OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is
5 indicative of metastasis if low level expression of an OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of
10 expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at
15 least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The OSNA or OSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with ovarian cancers or
20 other ovary related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of ovarian disorders.

Staging

25 The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and
30 the expression level of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels are determined in a biological sample from a patient whose stage of

cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations
5 of an OSNA or OSP to determine the stage of an ovarian cancer.

Monitoring

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient
10 to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for
15 expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an OSNA or OSP that are associated with ovarian cancer.

If increased expression of an OSNA or OSP is associated with metastasis,
20 treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or
25 failure to progress to a neoplastic lesion. If decreased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are
30 determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs of the invention can also be monitored by analyzing levels of expression of the OSNAs and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of OSG, an aberrant modification of OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an OSG. Methods to detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

25 Methods of Detecting Noncancerous Ovarian Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal

ovary control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general, if high expression relative to a control of an OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and
5 more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if
10 the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an OSNA and/or OSP
15 is associated with a particular noncancerous ovarian disease by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovary tissue. In another embodiment, one may determine whether an OSNA or OSP exhibits structural alterations in a particular noncancerous ovarian disease state by
20 obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovary tissue.

Methods for Identifying Ovary Tissue

25

In another aspect, the invention provides methods for identifying ovary tissue. These methods are particularly useful in, *e.g.*, forensic science, ovary cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a
30 sample is ovary tissue or has ovary tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovary tissue or having ovary tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding

that the sample comprises ovary tissue. In a preferred embodiment, the OSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 138 through 238, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 137, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an OSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an OSP is expressed. Determining whether a sample expresses an OSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 138 through 238, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two OSNAs and/or OSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovary tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovary tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new ovary tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Ovary Tissue

In another aspect, the invention provides methods for producing engineered ovary tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an OSNA or an OSG into the cells, and growing the cells under conditions in

-106-

which they exhibit one or more properties of ovary tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal ovary tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered ovary tissue or cells comprises one of these cell types. In another
5 embodiment, the engineered ovary tissue or cells comprises more than one ovary cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovary cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more OSPs are introduced into cells,
10 preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode OSPs having amino acid sequences selected from SEQ ID NO: 138 through 238, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 137, or hybridizing nucleic acids, allelic variants or parts
15 thereof. In another highly preferred embodiment, an OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial ovary tissue may be used to treat patients who have lost some or all of their ovary function.

20 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a
25 preferred embodiment, the pharmaceutical composition comprises an OSNA or part thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 137, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an
30 OSP or fragment thereof. In a more preferred embodiment, the OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 138 through 238, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of

-107-

the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to an OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 138 through 238, or an antibody that binds to a polypeptide
5 that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

10 Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed.
15 (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and
20 parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

25 Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and
30 tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

- 5 Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

- 10 Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

- Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.
- 15

- Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.
- 20

- Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.
- 25

- Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.
- 30

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

10 Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil
15 base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of
25 the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature
30 of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

-110-

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

5 For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration
10 of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

15 For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as
20 cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for
25 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the
30 present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

-111-

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

5 The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for
10 OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed
15 by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in
20 one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of
25 circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

30 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

5 Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be
10 administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,
15 conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

20 Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovary function. As used herein,
25 "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

30 The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further
5 described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.,* Doronin *et al., J. Virol.* 75: 3314-24 (2001).

10 In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an OSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an OSP are
15 administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.,* Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes an OSP having the amino acid sequence of SEQ ID NO: 138 through
20 238, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement
25 defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an OSP having the amino acid sequence of SEQ ID NO: 138 through 238, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an OSG
30 antisense nucleic acid, are administered to downregulate transcription and/or translation of an OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

5 Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are
10 incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9
15 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid
20 molecule encoding an OSP, preferably an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 137, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

25 *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant OSP defect.

30 Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune

response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

5 In a preferred embodiment, the polypeptide is an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 137, or a part, allelic variant, substantially similar or hybridizing nucleic acid
10 thereof.

Antibody, Agonist and Antagonist Administration

 In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is
15 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred
20 embodiment, the antibody specifically binds to an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 137, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

 The present invention also provides methods for identifying modulators which bind to an OSP or have a modulatory effect on the expression or activity of an OSP.
25 Modulators which decrease the expression or activity of OSP (antagonists) are believed to be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian
30 cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to an OSP are key candidates for

further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an OSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of OSP is
5 administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an OSP.

In other embodiments a pharmaceutical composition comprising an agonist of an
10 OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP comprising an amino acid sequence of SEQ ID NO: 138 through 238, or a fusion protein, allelic variant, homolog, analog or
15 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 137, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Ovary Tissue

20 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovary or to specific cells in the ovary. In a preferred embodiment, an anti-OSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovary tissue needs to be
25 selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovary cell function.

In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be
30 useful for determining and monitoring ovary function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

Example 1: Gene Expression analysis

OSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the OSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), and differential expression in cancer tissue (CLASP 5). cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within

-118-

each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

10 The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.
- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-137 are listed below:

SEQ ID NO: 1	DEX0257_1	CLASP2
30 SEQ ID NO: 2	DEX0257_2	CLASP2
SEQ ID NO: 3	DEX0257_3	CLASP2
SEQ ID NO: 4	DEX0257_4	CLASP5 CLASP1
SEQ ID NO: 5	DEX0257_5	CLASP5 CLASP1
SEQ ID NO: 6	DEX0257_6	CLASP2

-119-

	SEQ ID NO: 7	DEX0257_7	CLASP2
	SEQ ID NO: 8	DEX0257_8	CLASP2
	SEQ ID NO: 9	DEX0257_9	CLASP2
	SEQ ID NO: 10	DEX0257_10	CLASP2
5	SEQ ID NO: 11	DEX0257_11	CLASP2
	SEQ ID NO: 12	DEX0257_12	CLASP2
	SEQ ID NO: 13	DEX0257_13	CLASP2
	SEQ ID NO: 14	DEX0257_14	CLASP2
	SEQ ID NO: 15	DEX0257_15	CLASP2
10	SEQ ID NO: 16	DEX0257_16	CLASP2
	SEQ ID NO: 17	DEX0257_17	CLASP2
	SEQ ID NO: 18	DEX0257_18	CLASP2
	SEQ ID NO: 19	DEX0257_19	CLASP2
	SEQ ID NO: 20	DEX0257_20	CLASP2
15	SEQ ID NO: 21	DEX0257_21	CLASP2
	SEQ ID NO: 22	DEX0257_22	CLASP2
	SEQ ID NO: 23	DEX0257_23	CLASP2
	SEQ ID NO: 24	DEX0257_24	CLASP2
	SEQ ID NO: 25	DEX0257_25	CLASP2
20	SEQ ID NO: 26	DEX0257_26	CLASP2
	SEQ ID NO: 27	DEX0257_27	CLASP2
	SEQ ID NO: 28	DEX0257_28	CLASP2
	SEQ ID NO: 29	DEX0257_29	CLASP2
	SEQ ID NO: 30	DEX0257_30	CLASP2
25	SEQ ID NO: 31	DEX0257_31	CLASP2
	SEQ ID NO: 32	DEX0257_32	CLASP2
	SEQ ID NO: 33	DEX0257_33	CLASP2
	SEQ ID NO: 34	DEX0257_34	CLASP2
	SEQ ID NO: 35	DEX0257_35	CLASP2
30	SEQ ID NO: 36	DEX0257_36	CLASP2
	SEQ ID NO: 37	DEX0257_37	CLASP2
	SEQ ID NO: 38	DEX0257_38	CLASP2
	SEQ ID NO: 39	DEX0257_39	CLASP2
	SEQ ID NO: 40	DEX0257_40	CLASP2
35	SEQ ID NO: 41	DEX0257_41	CLASP2
	SEQ ID NO: 42	DEX0257_42	CLASP2
	SEQ ID NO: 43	DEX0257_43	CLASP2
	SEQ ID NO: 44	DEX0257_44	CLASP2
	SEQ ID NO: 45	DEX0257_45	CLASP2
40	SEQ ID NO: 48	DEX0257_48	CLASP2
	SEQ ID NO: 49	DEX0257_49	CLASP2
	SEQ ID NO: 50	DEX0257_50	CLASP2
	SEQ ID NO: 51	DEX0257_51	CLASP2
	SEQ ID NO: 52	DEX0257_52	CLASP2
45	SEQ ID NO: 53	DEX0257_53	CLASP2
	SEQ ID NO: 54	DEX0257_54	CLASP2
	SEQ ID NO: 55	DEX0257_55	CLASP2
	SEQ ID NO: 56	DEX0257_56	CLASP2

-120-

	SEQ ID NO: 57	DEX0257_57	CLASP2
	SEQ ID NO: 58	DEX0257_58	CLASP2
	SEQ ID NO: 59	DEX0257_59	CLASP2
	SEQ ID NO: 60	DEX0257_60	CLASP2
5	SEQ ID NO: 61	DEX0257_61	CLASP2
	SEQ ID NO: 62	DEX0257_62	CLASP2
	SEQ ID NO: 63	DEX0257_63	CLASP2
	SEQ ID NO: 64	DEX0257_64	CLASP2
	SEQ ID NO: 65	DEX0257_65	CLASP2
10	SEQ ID NO: 66	DEX0257_66	CLASP2
	SEQ ID NO: 67	DEX0257_67	CLASP2
	SEQ ID NO: 68	DEX0257_68	CLASP2
	SEQ ID NO: 69	DEX0257_69	CLASP2 CLASP1
	SEQ ID NO: 70	DEX0257_70	CLASP2
15	SEQ ID NO: 71	DEX0257_71	CLASP2
	SEQ ID NO: 72	DEX0257_72	CLASP2
	SEQ ID NO: 73	DEX0257_73	CLASP2
	SEQ ID NO: 74	DEX0257_74	CLASP2
	SEQ ID NO: 75	DEX0257_75	CLASP2
20	SEQ ID NO: 76	DEX0257_76	CLASP2
	SEQ ID NO: 78	DEX0257_78	CLASP5 CLASP1
	SEQ ID NO: 79	DEX0257_79	CLASP2
	SEQ ID NO: 80	DEX0257_80	CLASP2
	SEQ ID NO: 81	DEX0257_81	CLASP1
25	SEQ ID NO: 82	DEX0257_82	CLASP2
	SEQ ID NO: 83	DEX0257_83	CLASP2
	SEQ ID NO: 84	DEX0257_84	CLASP2
	SEQ ID NO: 85	DEX0257_85	CLASP2
	SEQ ID NO: 86	DEX0257_86	CLASP2
30	SEQ ID NO: 87	DEX0257_87	CLASP2
	SEQ ID NO: 88	DEX0257_88	CLASP2
	SEQ ID NO: 89	DEX0257_89	CLASP5 CLASP1
	SEQ ID NO: 90	DEX0257_90	CLASP5 CLASP1
	SEQ ID NO: 91	DEX0257_91	CLASP5 CLASP1
35	SEQ ID NO: 92	DEX0257_92	CLASP1
	SEQ ID NO: 93	DEX0257_93	CLASP2
	SEQ ID NO: 94	DEX0257_94	CLASP2
	SEQ ID NO: 95	DEX0257_95	CLASP2
	SEQ ID NO: 96	DEX0257_96	CLASP2
40	SEQ ID NO: 97	DEX0257_97	CLASP2
	SEQ ID NO: 98	DEX0257_98	CLASP2
	SEQ ID NO: 99	DEX0257_99	CLASP2
	SEQ ID NO: 100	DEX0257_100	CLASP2
	SEQ ID NO: 101	DEX0257_101	CLASP2
45	SEQ ID NO: 102	DEX0257_102	CLASP2
	SEQ ID NO: 103	DEX0257_103	CLASP2
	SEQ ID NO: 104	DEX0257_104	CLASP2
	SEQ ID NO: 105	DEX0257_105	CLASP2

-121-

	SEQ ID NO: 106	DEX0257_106	CLASP2
	SEQ ID NO: 107	DEX0257_107	CLASP2
	SEQ ID NO: 108	DEX0257_108	CLASP2
	SEQ ID NO: 109	DEX0257_109	CLASP2
5	SEQ ID NO: 110	DEX0257_110	CLASP2
	SEQ ID NO: 111	DEX0257_111	CLASP2
	SEQ ID NO: 112	DEX0257_112	CLASP5 CLASP1
	SEQ ID NO: 113	DEX0257_113	CLASP5 CLASP1
	SEQ ID NO: 114	DEX0257_114	CLASP5 CLASP1
10	SEQ ID NO: 115	DEX0257_115	CLASP5 CLASP1
	SEQ ID NO: 117	DEX0257_117	CLASP2
	SEQ ID NO: 118	DEX0257_118	CLASP2
	SEQ ID NO: 119	DEX0257_119	CLASP2
	SEQ ID NO: 120	DEX0257_120	CLASP2
15	SEQ ID NO: 121	DEX0257_121	CLASP2
	SEQ ID NO: 122	DEX0257_122	CLASP2 CLASP1
	SEQ ID NO: 123	DEX0257_123	CLASP2 CLASP1
	SEQ ID NO: 124	DEX0257_124	CLASP2
	SEQ ID NO: 125	DEX0257_125	CLASP1
20	SEQ ID NO: 126	DEX0257_126	CLASP1
	SEQ ID NO: 127	DEX0257_127	CLASP2
	SEQ ID NO: 128	DEX0257_128	CLASP2
	SEQ ID NO: 129	DEX0257_129	CLASP1
	SEQ ID NO: 130	DEX0257_130	CLASP1
25	SEQ ID NO: 131	DEX0257_131	CLASP1
	SEQ ID NO: 132	DEX0257_132	CLASP2
	SEQ ID NO: 133	DEX0257_133	CLASP2
	SEQ ID NO: 134	DEX0257_134	CLASP2
	SEQ ID NO: 135	DEX0257_135	CLASP2
30	SEQ ID NO: 136	DEX0257_136	CLASP2
	SEQ ID NO: 137	DEX0257_137	CLASP2

CLASP Expression percentage levels for DEX0257 genes

35	SEQ ID NO: 1	OVR .0051				
	SEQ ID NO: 2	OVR .0064				
	SEQ ID NO: 3	OVR .0064				
	SEQ ID NO: 4	OVR .0032	BRN .0003	UTR .0004	KID .0006	STO .0021
	SEQ ID NO: 5	OVR .0032	BRN .0003	UTR .0004	KID .0006	STO .0021
40	SEQ ID NO: 6	OVR .0023				
	SEQ ID NO: 7	OVR .0023				
	SEQ ID NO: 8	OVR .0023				
	SEQ ID NO: 9	OVR .0023				
	SEQ ID NO: 10	OVR .0023				
45	SEQ ID NO: 11	OVR .0023				
	SEQ ID NO: 12	OVR .0023				
	SEQ ID NO: 13	OVR .0023				
	SEQ ID NO: 14	OVR .0023				

-122-

	SEQ ID NO: 15	OVR .0023	
	SEQ ID NO: 16	OVR .0023	
	SEQ ID NO: 17	OVR .0023	LIV .0024
	SEQ ID NO: 18	OVR .0023	LIV .0024
5	SEQ ID NO: 19	OVR .0063	
	SEQ ID NO: 20	OVR .0063	
	SEQ ID NO: 21	OVR .0063	
	SEQ ID NO: 22	OVR .0063	
	SEQ ID NO: 23	OVR .0056	
10	SEQ ID NO: 24	OVR .0056	
	SEQ ID NO: 25	OVR .0056	
	SEQ ID NO: 26	OVR .0056	
	SEQ ID NO: 27	OVR .0059	
	SEQ ID NO: 28	OVR .0059	
15	SEQ ID NO: 29	OVR .0059	
	SEQ ID NO: 30	OVR .0059	
	SEQ ID NO: 31	OVR .0059	
	SEQ ID NO: 32	OVR .0059	
	SEQ ID NO: 33	OVR .0051	
20	SEQ ID NO: 34	OVR .0051	BRN .0022
	SEQ ID NO: 35	OVR .0051	
	SEQ ID NO: 36	OVR .0051	
	SEQ ID NO: 37	OVR .0051	
	SEQ ID NO: 38	OVR .0051	
25	SEQ ID NO: 39	OVR .0051	
	SEQ ID NO: 40	OVR .0051	
	SEQ ID NO: 41	OVR .0051	
	SEQ ID NO: 42	OVR .0051	
	SEQ ID NO: 43	OVR .0051	
30	SEQ ID NO: 44	OVR .0051	
	SEQ ID NO: 45	OVR .0051	
	SEQ ID NO: 48	OVR .0051	
	SEQ ID NO: 49	OVR .0051	
	SEQ ID NO: 50	OVR .0051	
35	SEQ ID NO: 51	OVR .0051	
	SEQ ID NO: 52	OVR .0051	
	SEQ ID NO: 53	OVR .0051	
	SEQ ID NO: 54	OVR .0062	
	SEQ ID NO: 55	OVR .0062	
40	SEQ ID NO: 56	OVR .0062	
	SEQ ID NO: 57	OVR .0062	
	SEQ ID NO: 58	OVR .0062	
	SEQ ID NO: 59	OVR .0062	
	SEQ ID NO: 60	OVR .0062	
45	SEQ ID NO: 61	OVR .0062	
	SEQ ID NO: 62	OVR .0062	
	SEQ ID NO: 63	OVR .0062	
	SEQ ID NO: 64	OVR .0062	

-123-

	SEQ ID NO: 65	OVR .0062				
	SEQ ID NO: 66	OVR .0062				
	SEQ ID NO: 67	OVR .0062	BRN .0004			
	SEQ ID NO: 68	OVR .0062	BRN .0004			
5	SEQ ID NO: 69	OVR .0088				
	SEQ ID NO: 70	OVR .0062				
	SEQ ID NO: 71	OVR .0062				
	SEQ ID NO: 72	OVR .0062				
	SEQ ID NO: 73	OVR .0062				
10	SEQ ID NO: 74	OVR .0062				
	SEQ ID NO: 75	OVR .0062				
	SEQ ID NO: 76	OVR .0062				
	SEQ ID NO: 78	OVR .0032	CON .0007	PRO .0007	CRD .002	CRD .0023
	SEQ ID NO: 79	OVR .0059				
15	SEQ ID NO: 80	OVR .0131				
	SEQ ID NO: 81	OVR .0032	FTS .0001	BRN .0003	KID .0006	NRV .0009
	SEQ ID NO: 82	OVR .0042	PRO .0019	THR .0127		
	SEQ ID NO: 83	OVR .0042	PRO .0019	THR .0127		
	SEQ ID NO: 84	OVR .0023				
20	SEQ ID NO: 85	OVR .0023				
	SEQ ID NO: 86	OVR .0062				
	SEQ ID NO: 87	OVR .0062				
	SEQ ID NO: 88	OVR .0051	CON .0007			
	SEQ ID NO: 89	OVR .0043				
25	SEQ ID NO: 90	OVR .0043				
	SEQ ID NO: 91	OVR .0032	FTS .0003	INL .0004	INS .001	
	SEQ ID NO: 92	OVR .0032	INL .0012	KID .0013	TNS .0017	CRD .002
	SEQ ID NO: 93	OVR .0062				
	SEQ ID NO: 94	OVR .0052				
30	SEQ ID NO: 95	OVR .0064	STO .0185			
	SEQ ID NO: 96	OVR .0064	STO .0185			
	SEQ ID NO: 97	OVR .0097				
	SEQ ID NO: 98	OVR .0097				
	SEQ ID NO: 99	OVR .0052				
35	SEQ ID NO: 100	OVR .0064				
	SEQ ID NO: 101	OVR .0058	MAM .002			
	SEQ ID NO: 102	OVR .0058	MAM .002			
	SEQ ID NO: 103	OVR .0043				
	SEQ ID NO: 104	OVR .0043				
40	SEQ ID NO: 105	OVR .0023				
	SEQ ID NO: 106	OVR .0052				
	SEQ ID NO: 107	OVR .0052				
	SEQ ID NO: 108	OVR .0062				
	SEQ ID NO: 109	OVR .0064				
45	SEQ ID NO: 110	OVR .0051	CON .0007	UTR .005		
	SEQ ID NO: 111	OVR .0051	CON .0007	UTR .005		
	SEQ ID NO: 112	OVR .0064	FTS .0003	LNG .0004	LNG .0006	BLO .0006
	SEQ ID NO: 113	OVR .0064	FTS .0003	LNG .0004	LNG .0006	BLO .0006

-124-

SEQ ID NO: 114 OVR .0064 FTS .0003 LNG .0004 LNG .0006 BLO .0006
 SEQ ID NO: 115 OVR .0064 FTS .0003 LNG .0004 LNG .0006 BLO .0006
 SEQ ID NO: 117 OVR .0064
 SEQ ID NO: 118 OVR .0023
 5 SEQ ID NO: 119 OVR .0023
 SEQ ID NO: 120 OVR .0064
 SEQ ID NO: 121 OVR .0064
 SEQ ID NO: 122 OVR .0034
 SEQ ID NO: 123 OVR .0034
 10 SEQ ID NO: 124 OVR .0023
 SEQ ID NO: 125 OVR .0021 MSL .002
 SEQ ID NO: 126 OVR .0021 MSL .002
 SEQ ID NO: 127 OVR .0093
 SEQ ID NO: 128 OVR .0093
 15 SEQ ID NO: 129 OVR .0063 FTS .0003 LNG .0004 INL .0004 CON .0007
 SEQ ID NO: 130 OVR .0063 FTS .0003 LNG .0004 INL .0004 CON .0007
 SEQ ID NO: 131 OVR .0063 FTS .0003 LNG .0004 INL .0004 CON .0007
 SEQ ID NO: 132 OVR .0052
 SEQ ID NO: 133 OVR .0052
 20 SEQ ID NO: 134 OVR .0063
 SEQ ID NO: 135 OVR .0063
 SEQ ID NO: 136 OVR .0063
 SEQ ID NO: 137 OVR .0063
 25 Abbreviation for tissues:
 BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine,
 Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL
 Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR
 Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus
 30

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation
 detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The
 method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5'
 35 reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity
 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected
 by the laser detector of the Model 7700 Sequence Detection System (PE Applied
 Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to
 standardize the amount of sample RNA added to the reaction and normalize for Reverse
 40 Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate
 dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this
 endogenous control. To calculate relative quantitation between all the samples studied,

-125-

the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

- 5 The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

- One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

- The relative levels of expression of the OSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

- In the analysis of matching samples, the OSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

- Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

- 30 Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 137 being a diagnostic marker for cancer.

-126-

Sequence
Dex0097_24 (ovr125-sqovr007)

Sequence ID No.
DEX0257_33 (SEQ ID No. 33)

Semi-quantitative PCR was done using the following primers:

5

Primer	DexSeqID	From	To	Primer Length
sqovr007F	DEX0257_33	15	37	23
sqovr007R	DEX0257_33	233	213	21

Data from the semiQ-PCR experiment showed that sqovr007 was overexpressed in 3 of 6 (50%) ovarian cancer matching samples. Sqovr007 was advanced to quantitative PCR and named ovr125.

10

Quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
ovr125F	DEX0257_33	17	38	22
ovr125R	DEX0257_33	120	98	23
ovr125probe	DEX0257_33	47	76	30

Table 1. The absolute numbers are relative levels of expression of ovr125 in 24 normal samples from 24 different tissues. All the values are compared to normal brain (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

Tissue	Normal
Adrenal Gland	0.00
Bladder	0.00
Brain	1.00
Cervix	0.00
Colon	0.15
Endometrium	0.00
Esophagus	0.00
Heart	0.00
Kidney	0.61
Liver	0.00
Lung	0.00
Mammary	0.68
Muscle	0.08
Ovary	7.73
Pancreas	2.59
Prostate	0.00
Rectum	0.00
Small Intestine	0.00
Spleen	12.47
Stomach	0.00

15

-127-

Testis	0.00
Thymus	9.09
Trachea	1.74
Uterus	0.00

The relative levels of expression in the table above show that ovr125 mRNA expression is detected in the pool of normal spleen, thymus followed by ovary. Fourteen normal samples do not show expression of ovr125.

- 5 The absolute numbers in the table were obtained analyzing pools of samples of a particular tissue from different individuals. They cannot be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in the table below.

- 10 The relative levels of expression of ovr125 in 48 pairs of matching samples were analyzed. All the values are compared to normal brain (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. In addition, 9 unmatched cancer samples (from ovary) and 7 unmatched normal samples (from ovary) were also tested.

15

Sample ID	Tissue	Cancer	Normal Adjacent Tissue	Normal
OvrA084	ovary 1	72.00	59.30	
OvrG021	ovary 2	36.76	91.46	
Ovr3710	ovary 3	48.17		
Ovr638A	ovary 4	157.59		
Ovr63A	ovary 5	163.14		
Ovr773O	ovary 6	11.71		
Ovr988Z	ovary 7	38.19		
Ovr1005O	ovary 8	54.19		
Ovr1028	ovary 9	168.31		
Ovr1040O	ovary 10	38.72		
Ovr105O	ovary 11	56.69		
OvrC087	ovary 12			8.40
OvrC109	ovary 13			14.88
Ovr18GA	ovary 14			203.66
Ovr206I	ovary 15			341.32
Ovr20GA	ovary 16			58.49
Ovr247A	ovary 17			85.63
Ovr25GA	ovary 18			496.28
Bld46XK	bladder 1	0.00	0.00	
BldTR14	bladder 2	82.14	79.89	
Liv15XA	liver 1	2.35	0.00	
Utr135XO	uterus 1	115.36	121.94	
Tst647T	testis 1	16.56	128.89	

-128-

ClnDC63	Colon 1	82.14	27.57	
Thr590D	thymus 1	25.11	7.52	
LngSQ80	lung 1	66.72	11.27	
Endo12XA	endometrium 1	71.01	0.00	
Mam986	mammary gland 1	0.00	0.00	

0.00= Negative

The table above represents 40 samples in 10 different tissues. The two tables above represent a combined total of 64 samples in 24 human tissue types.

- Comparisons of the level of mRNA expression in ovarian cancer samples with normal ovarian tissue are shown. The analysis of two ovarian matching samples showed no difference (ovary 1) or downregulation (ovary 2) when cancer was compared with normal adjacent tissue. For the unmatched ovarian samples, the median of the normal ovarian samples (85.63) was compared with the cancer samples. Three out of nine ovarian cancer samples (ovary 4, 5, and 9: 33%) showed expression over 1.5 times the value of the median for normal ovary.

Sequence**Sequence ID #****Dex0097_29 (sqovr008)****DEX0257_39 (SEQ ID NO:39)**

Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqovr008F	DEX0257_39	62	83	22
sqovr008R	DEX0257_39	195	174	22

- Table 1.** The relative levels of expression of sqovr008 in 12 normal samples from 12 different tissues were analyzed. These RNA samples are from single individual or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	NORMAL
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0

-129-

Small Intestine	0
Stomach	0
Testis	0
Uterus	0

Relative levels of expression in the table above show no sqovr008 expression in any of the normal tissues analyzed.

The relative levels of expression of sqovr008 in 12 cancer samples from 12
5 different tissues were analyzed. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

10

TISSUE	CANCER
Bladder	0
Breast	0
Colon	0
Kidney	0
Liver	0
Lung	1
Ovary	0
Pancreas	0
Prostate	0
Stomach	0
Testes	1
Uterus	0

Relative levels of expression in the table above show that sqovr008 is expressed only in lung and testes carcinomas.

The relative levels of expression of sqovr008 in 6 ovarian cancer matching
15 samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were
analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of
20 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

-130-

SAMPLE ID	TISSUE	CANCER	NORMAL ADJACENT TISSUE
VNM-00116D04/N05	ovary 1	100	0
VNM-00291D01/N04	ovary 2	0	0
S99-5693A/B	ovary 3	1	0
9708G021SP1/N1	ovary 4	0	0
9704A081F/2D	ovary 5	0	0
9803G010SP1/N1	ovary 6	0	1

Relative levels of expression in Table 2 shows that sqovr008 is upregulated in 2 out of 6 (33%) of the matching samples analyzed.

Experiments are underway to design and test primers and probe for quantitative
5 PCR.

Sequence	Sequence ID #
Dex0097_74 (sqovr013)	DEX0257_103 (SEQ ID NO: 103)
	DEX0257_104 (SEQ ID NO: 104)

Semi-quantitative PCR was done using the following primers:

Primer	DexSeqID	From	To	Primer Length
sqovr013F	DEX0257_104	1538	1514	25
sqovr013F	DEX0257_103	17	41	25
sqovr013R	DEX0257_103	163	139	25
sqovr013R	DEX0257_104	1392	1416	25

15 The relative levels of expression of sqovr013 in 12 normal samples from 12 different tissues were analyzed. These RNA samples are from single individual or are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

Using Polymerase Chain Reaction (PCR) technology expression levels were
20 analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

TISSUE	NORMAL
Breast	10
Colon	10
Endometrium	10
Kidney	100
Liver	10
Lung	0
Ovary	100

-131-

Prostate	10
Small Intestine	100
Stomach	100
Testis	10
Uterus	0

Relative levels of expression in Table 1 show sqovr013 expression in most of the normal tissues analyzed, including ovary among the tissues with highest expression.

The relative levels of expression of sqovr013 in 12 cancer samples from 12
5 different tissues were analyzed. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

10

TISSUE	CANCER
Bladder	1000
Breast	1
Colon	10
Kidney	10
Liver	10
Lung	10
Ovary	10
Pancreas	100
Prostate	100
Stomach	1
Testes	1
Uterus	10

Relative levels of expression in the table above show that sqovr013 is expressed in all carcinomas tested with highest expression in bladder carcinoma.

The relative levels of expression of sqovr013 in 6 ovarian cancer matching
15 samples were analyzed. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression.
20 A positive reaction in the most dilute sample indicates the highest relative expression value.

SAMPLE ID	TISSUE	CANCER	NORMAL ADJACENT TISSUE
VNM-00116D04/N05	ovary 1	100	1
VNM-00291D01/N04	ovary 2	100	1
S99-5693A/B	ovary 3	100	10
9708G021SP1/N1	ovary 4	1	10
9704A081F/2D	ovary 5	10	10
9803G010SP1/N1	ovary 6	1	1

Relative levels of expression in Table 2 shows that sqovr013 is upregulated in 3 out of 6 (50%) of the matching samples analyzed.

- 5 Experiments are underway to design and test primers and probe for quantitative PCR.

Example 3: Protein Expression

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

15 An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. OSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using 25 primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI

cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without
5 a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

10 **Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine
15 serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from
20 the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.
25 Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The
30 splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such

-134-

antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by

5 reference).

	DEX0257_140	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	19-29	1.09	11
	414-425	1.01	12
10	DEX0257_149	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	33-45	1.18	13
	17-26	1.00	10
	DEX0257_155	Antigenicity Index(Jameson-Wolf)	
15	positions	AI	avg length
	20-35	1.05	16
	DEX0257_157	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	60-70	1.29	11
20	14-57	1.11	44
	DEX0257_160	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	26-36	1.31	11
	DEX0257_161	Antigenicity Index(Jameson-Wolf)	
25	positions	AI	avg length
	17-50	1.12	34
	DEX0257_163	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	4-17	1.14	14
30	DEX0257_166	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	9-18	1.05	10
	DEX0257_167	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
35	37-53	1.01	17
	DEX0257_172	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	40-52	1.03	13
	DEX0257_180	Antigenicity Index(Jameson-Wolf)	
40	positions	AI	avg length
	42-58	1.42	17
	18-38	1.10	21
	DEX0257_200	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
45	47-63	1.09	17
	DEX0257_201	Antigenicity Index(Jameson-Wolf)	

-135-

	positions	AI	avg length
	103-113	1.18	11
	67-87	1.11	21
	DEX0257_207	Antigenicity Index(Jameson-Wolf)	
5	positions	AI	avg length
	29-38	1.08	10
	DEX0257_209	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	13-24	1.11	12
10	DEX0257_212	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	264-279	1.35	16
	151-171	1.19	21
	361-374	1.08	14
15	333-344	1.02	12
	DEX0257_218	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	7-37	1.16	31
	DEX0257_220	Antigenicity Index(Jameson-Wolf)	
20	positions	AI	avg length
	2-14	1.22	13
	33-44	1.21	12
	DEX0257_221	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
25	89-104	1.16	16
	19-58	1.14	40
	136-165	1.12	30
	115-130	1.11	16
	359-370	1.08	12
30	DEX0257_225	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	25-34	1.19	10
	DEX0257_231	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
35	448-476	1.20	29
	246-277	1.19	32
	868-888	1.19	21
	532-631	1.17	100
	45-54	1.10	10
40	817-833	1.09	17
	314-382	1.08	69
	784-811	1.06	28
	387-423	1.04	37
	425-440	1.04	16
45	225-240	1.03	16
	638-675	1.01	38
	838-865	1.01	28
	DEX0257_233	Antigenicity Index(Jameson-Wolf)	

-136-

	positions	AI	avg length
	348-385	1.14	38
	48-74	1.12	27
	230-252	1.06	23
5	322-342	1.01	21
	DEX0257_235	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	7-16	1.28	10

10 Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs

15 of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

20	DEX0257_140	Amidation 202-205; Asn_Glycosylation 185-188; Camp_Phospho_Site 405-408; Ck2_Phospho_Site 75-78;159-162;223-226;267-270;291-294;414-417; Myristyl 9-14;91-96;155-160;262-267;268-273;340-345;389-394;435-440; Pkc_Phospho_Site 159-161;191-193;254-256;392-394;393-395;
	DEX0257_142	Ck2_Phospho_Site 18-21; Pkc_Phospho_Site 37-39; Tyr_Phospho_Site 4-11;
25	DEX0257_144	Ck2_Phospho_Site 21-24;
	DEX0257_146	Myristyl 18-23;23-28;44-49;47-52;73-78;92-97; Prokar_Lipoprotein 39-49;
	DEX0257_148	Myristyl 19-24;84-89; Prokar_Lipoprotein 61-71;
	DEX0257_149	Amidation 59-62; Ck2_Phospho_Site 23-26; Myristyl 13-18;41-46;
30	DEX0257_150	Ck2_Phospho_Site 4-7; Pkc_Phospho_Site 19-21;
	DEX0257_154	Ck2_Phospho_Site 16-19;
	DEX0257_156	Ck2_Phospho_Site 16-19;
	DEX0257_157	Amidation 19-22; Ck2_Phospho_Site 47-50; Myristyl 14-19;15-20; Pkc_Phospho_Site 5-7;55-57;
35	DEX0257_158	Myristyl 59-64; Pkc_Phospho_Site 26-28;
	DEX0257_159	Ck2_Phospho_Site 15-18; Pkc_Phospho_Site 3-5;
	DEX0257_160	Myristyl 30-35;
	DEX0257_161	Camp_Phospho_Site 24-27; Pkc_Phospho_Site 31-33;
40	DEX0257_162	Myristyl 2-7; Pkc_Phospho_Site 3-5;
	DEX0257_163	Pkc_Phospho_Site 14-16;
	DEX0257_164	Pkc_Phospho_Site 57-59;
	DEX0257_165	Asn_Glycosylation 44-47; Ck2_Phospho_Site 37-40;51-54; Pkc_Phospho_Site 50-52;

-137-

	DEX0257_167	Ck2_Phospho_Site 71-74;96-99; Myristyl 66-71;67-72;68-73; Pkc_Phospho_Site 71-73;96-98;
	DEX0257_169	Ck2_Phospho_Site 38-41;
	DEX0257_170	Ck2_Phospho_Site 25-28; Myristyl 18-23;
5	DEX0257_171	Myristyl 37-42; Pkc_Phospho_Site 38-40; Tyr_Phospho_Site 13-19;
	DEX0257_172	Ck2_Phospho_Site 12-15; Pkc_Phospho_Site 2-4;49-51;
	DEX0257_174	Asn_Glycosylation 7-10; Ck2_Phospho_Site 9-12; Myristyl 5-10;
	DEX0257_175	Camp_Phospho_Site 53-56; Ck2_Phospho_Site 39-42;41-44;
10	DEX0257_176	Myristyl 12-17;15-20;16-21;20-25;22-27;59-64;
	DEX0257_178	Camp_Phospho_Site 11-14; Pkc_Phospho_Site 14-16;
	DEX0257_180	Pkc_Phospho_Site 25-27;
	DEX0257_181	Myristyl 5-10;
15	DEX0257_183	Myristyl 4-9;
	DEX0257_184	Ck2_Phospho_Site 7-10; Pkc_Phospho_Site 19-21;
	DEX0257_186	Amidation 21-24; Camp_Phospho_Site 23-26; Ck2_Phospho_Site 12-15; Myristyl 41-46;44-49;
	DEX0257_187	Ck2_Phospho_Site 11-14;
20	DEX0257_188	Ck2_Phospho_Site 46-49; Myristyl 97-102;
	DEX0257_190	Myristyl 15-20;
	DEX0257_191	Myristyl 29-34; Pkc_Phospho_Site 35-37;
	DEX0257_192	Pkc_Phospho_Site 27-29; Rgd 30-32;
25	DEX0257_193	Asn_Glycosylation 19-22; Ck2_Phospho_Site 21-24;
	DEX0257_194	Pkc_Phospho_Site 26-28;
	DEX0257_197	Camp_Phospho_Site 117-120; Ck2_Phospho_Site 78-81; Myristyl 17-22;98-103; Pkc_Phospho_Site 22-24;109-111;115-117;116-118;120-122;
30	DEX0257_198	Asn_Glycosylation 14-17; Pkc_Phospho_Site 13-15;
	DEX0257_199	Asn_Glycosylation 17-20;
	DEX0257_200	Myristyl 2-7;6-11;
	DEX0257_201	Asn_Glycosylation 25-28; Ck2_Phospho_Site 37-40;
35	DEX0257_203	Camp_Phospho_Site 49-52; Ck2_Phospho_Site 32-35;
	DEX0257_204	Pkc_Phospho_Site 22-24;
	DEX0257_205	Asn_Glycosylation 11-14;108-111;127-130; Ck2_Phospho_Site 28-31; Myristyl 78-83; Pkc_Phospho_Site 13-15;74-76;82-84;
40	DEX0257_206	Asn_Glycosylation 55-58; Pkc_Phospho_Site 39-41;
	DEX0257_207	Ck2_Phospho_Site 28-31; Myristyl 21-26; Pkc_Phospho_Site 28-30;
	DEX0257_210	Asn_Glycosylation 30-33; Myristyl 31-36;100-105;103-108;
45	DEX0257_211	Pkc_Phospho_Site 23-25;
	DEX0257_212	Asn_Glycosylation 9-12; Pkc_Phospho_Site 4-6;
	DEX0257_212	Asn_Glycosylation 9-12;24-27;64-67; Ck2_Phospho_Site 49-52;
	DEX0257_212	Myristyl 41-46;
	DEX0257_212	Ck2_Phospho_Site 21-24;
	DEX0257_212	Pkc_Phospho_Site 16-18;
	DEX0257_212	Asn_Glycosylation 43-46;69-72;93-96;303-306;368-371;462-465;
	DEX0257_212	Camp_Phospho_Site 360-363; Ck2_Phospho_Site 272-275;284-287;288-291;466-469; Myristyl 76-81; Pkc_Phospho_Site 45-

		47;64-66;96-98;163-165;206-208;236-238;293-295;294-296;339-341;359-361;363-365;370-372; Tyr_Phospho_Site 164-171;165-171;
	DEX0257_213	Camp_Phospho_Site 22-25;29-32; Pkc_Phospho_Site 32-34;
5	DEX0257_214	Asn_Glycosylation 36-39;
	DEX0257_216	Ck2_Phospho_Site 26-29;
	DEX0257_217	Ck2_Phospho_Site 19-22;
	DEX0257_218	Amidation 43-46;50-53; Camp_Phospho_Site 11-14; Myristyl 9-14; Pkc_Phospho_Site 10-12;50-52;
10	DEX0257_220	Myristyl 44-49;60-65; Pkc_Phospho_Site 34-36;73-75;
	DEX0257_221	Asn_Glycosylation 105-108;201-204; Camp_Phospho_Site 73-76; Ck2_Phospho_Site 4-7;23-26;44-47;107-110;359-362;372-375; Fork_Head_1 54-67; Fork_Head_2 98-104; Myristyl 37-42;38-43;39-44;40-45;125-130;165-170;168-173;170-175;171-176;175-180;177-182;237-242;269-274;278-283;342-347;368-373; Pkc_Phospho_Site 20-22;23-25;101-103; Prokar_Lipoprotein 166-176;
15		
	DEX0257_223	Asn_Glycosylation 21-24;
	DEX0257_224	Myristyl 26-31;
20	DEX0257_225	Asn_Glycosylation 28-31;46-49; Myristyl 2-7;
	DEX0257_227	Myristyl 18-23;46-51; Pkc_Phospho_Site 11-13;
	DEX0257_228	Asn_Glycosylation 14-17; Myristyl 11-16; Pkc_Phospho_Site 16-18;27-29;80-82;
	DEX0257_229	Asn_Glycosylation 70-73;87-90; Camp_Phospho_Site 19-22;
25		Ck2_Phospho_Site 22-25;72-75;79-82; Myristyl 3-8;7-12;10-15; Pkc_Phospho_Site 53-55;79-81;
	DEX0257_230	Asn_Glycosylation 23-26; Camp_Phospho_Site 62-65; Pkc_Phospho_Site 27-29;61-63;
	DEX0257_231	Amidation 709-712; Asn_Glycosylation 193-196;213-216;220-223;781-784;908-911; Camp_Phospho_Site 112-115;361-364; Ck2_Phospho_Site 4-7;13-16;97-100;162-165;363-366;503-506;633-636; Cytochrome_C 772-777; Myristyl 52-57;304-309;429-434;734-739; Pkc_Phospho_Site 4-6;23-25;45-47;46-48;97-99;172-174;176-178;215-217;293-295;360-362;367-369;405-407;416-418;433-435;507-509;554-556;563-565;584-586;612-614;629-631;696-698;797-799;881-883;892-894; Zinc_Finger_C2h2 240-260;268-288;296-316;324-344;352-372;380-400;408-428;436-456;464-484;520-540;548-568;576-596;604-624;632-652;660-680;688-708;716-736;744-764;800-820;828-848;884-904;
30		
	DEX0257_232	Ck2_Phospho_Site 93-96;101-104; Myristyl 27-32;115-120;118-123;122-127;125-130;133-138;146-151;152-157;156-161;170-175;175-180;270-275;274-279;276-281;317-322; Pkc_Phospho_Site 28-30;194-196;
35		
	DEX0257_233	Amidation 27-30; Asn_Glycosylation 250-253;450-453; Bpti_Kunitz 345-363; Ck2_Phospho_Site 51-54;152-155;415-418;452-455; Myristyl 14-19;58-63;97-102;213-218;224-229;235-240;240-245;340-345;348-353;349-354;352-357;372-377;478-
40		
45		

-139-

- 483; Pkc_Phospho_Site 104-106;218-220;409-411;481-483;
 Tyr_Phospho_Site 208-215;
 DEX0257_234 Ck2_Phospho_Site 66-69; Myristyl 79-84;83-88;
 Pkc_Phospho_Site 56-58;
 5 DEX0257_235 Pkc_Phospho_Site 13-15;
 DEX0257_236 Ck2_Phospho_Site 3-6;
 DEX0257_237 Pkc_Phospho_Site 19-21;
 DEX0257_238 Asn_Glycosylation 79-82; Camp_Phospho_Site 40-43;
 Ck2_Phospho_Site 45-48;

10

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

- RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using
 15 protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 137. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*,
 20 *Science* 278(5340): 1054-9 (1997).

- PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then
 25 cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

- Genomic rearrangements may also be determined. Genomic clones are
 30 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

- Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium
 35 iodide, producing a combination of C-and R-bands. Aligned images for precise mapping

are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical
5 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological
10 Sample**

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells
15 are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate,
20 at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

25 The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater.

-142-

Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein
5 et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent
10 cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable
15 carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with
20 liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as
25 liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic
30 acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid,

aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

5 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

10 Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

15 Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

20 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

30 It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a

pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 µg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf

-145-

intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

30 Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the

introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide
5 by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al.
10 (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

15 The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) *Ann. NY Acad. Sci.* 772: 126-139 and Abdallah B. et al. (1995) *Biol. Cell* 85
20 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the
25 art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

30 The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach,

intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue
5 ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are
10 differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or
15 RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be
20 determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the
25 nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard
30 recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

-148-

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell.

Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campbell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be

accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples
5 obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or
10 crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous
15 transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

20 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

25 **Example 14: Knock-Out Animals**

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by
30 reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions

of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via
5 targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the
10 recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a
15 patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or
20 alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

25 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

30 Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or

-152-

vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

5 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

10 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

15 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments,
20 which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 138 through 238;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 137;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of an ovary specific nucleic acid
(OSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to an
ovary specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to an OSNA in the
sample, wherein the detection of the hybridization indicates the presence of an OSNA in
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60%
15 sequence identity to of SEQ ID NO: 138 through 238; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 137.

12. An antibody or fragment thereof that specifically binds to the polypeptide
20 according to claim 11.

13. A method for determining the presence of an ovary specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under
25 conditions in which the antibody will selectively bind to the ovary specific protein; and

(b) detecting binding of the antibody to an ovary specific protein in the sample, wherein the detection of binding indicates the presence of an ovary specific protein in the sample.

30 14. A method for diagnosing and monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:

-155-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovary specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of ovarian cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with ovarian cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

SEQUENCE LISTING

<110> Sun, Yongming
 Recipon, Herve
 Salceda, Susana
 Liu, Chenghua
 diaDexus, Inc.

<120> Compositions and Methods Relating to Ovary Specific
 Genes and Proteins

<130> DEX-0258

<140>

<141>

<150> 60/246,640

<151> 2000-11-08

<160> 238

<170> PatentIn Ver. 2.1

<210> 1

<211> 151

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (77)..(78)

<223> a, c, g or t

<220>

<221> unsure

<222> (82)

<223> a, c, g or t

<220>

<221> unsure

<222> (129)

<223> a, c, g or t

<220>

<221> unsure

<222> (143)

<223> a, c, g or t

<400> 1

atcccactca atcatggggt attatgttcc tgatgtgctg ttagatttgg tttctagta 60
 ttggctgag gatttttngc anttaatgtt ccactagagg taattgccct ggttggtggc 120
 ttgtctggng ttgggatcag ganttatgct g 151

<210> 2

<211> 59

<212> DNA

<213> Homo sapiens

<400> 2

cgataaatgt agatatgaaa gcagactaca gtataaaaca ctgctcagaa acattttgc 59

<210> 3
 <211> 2330
 <212> DNA
 <213> Homo sapiens

<400> 3
 taaaacacct ggcccagggtg gtacaaacac caactcgggg accacgagggc taagccccca 60
 aaggcgggga ccaagagcag agccatggta ccgggaaagg ttcaacaaaa ataaggtaaa 120
 aaaatgtgtg tgtctgtgaa aaaatttggc ccgttacggg ggctaaaaatg caatgggtgca 180
 acctcagctg atggccaagt cggctaccta gggtcaaacg aatctaccgc gctaagcctc 240
 cgaagtagct ggtaatacag ggcatgcgtc acgatgccc gccaattgtg catgcttagt 300
 acagactggg ttccaccatg ttggtcagcc cctcgaactc ctgacctcag gtgattcgct 360
 gcctcggcat cccaaagtgc tgggatacag gcctgagcca ctgcgcaccg gtgcgaaaaat 420
 gtttctgagc agtggtttat actgtagtct gctttcatat ctacatttat cgtatttata 480
 tattttattta ctgagacagg gtctcactct gtttcccagg ctggagtatg gtggcacaat 540
 cttggccttac tgcaacctcc acctcccagg ctcaagcaat cctccacact cagcctocca 600
 agtagctggg actagagatg tatgccatca cactcggtt tgtgtgtgtg tgtgtgtgtg 660
 tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgaggtttca ctatgtttcc caggctggct 720
 tcgaactcct aagctcaagc gatccacca tctcggcctc ccaaagtgtc gggattatag 780
 gcataagtca ctgcacctgg ccattggcatg attcttttat ctctcctggg gctgaactcc 840
 ctacatttgc ttacacctgg cctggaagac ccaagatccc cctcacaata ctacttctat 900
 acccagggcc aggtgatgtg ttggttagt tcaggactga gctttattat gcattctcca 960
 gcaggcaacc tggggcctct gatactgcc gggagagctg ggggaatgga gctgttcctg 1020
 acttcttaca caaggagtgg ctaatcttct gtcttttttc taaccaaagc catctctgga 1080
 ccaccaagag caagtgggca gaggtccctc atcctggcag gagagcagag ctgccagcaa 1140
 tgaaagaaca gaaataggca aatgagaact caggctctgt cacagaaccc tctcctctct 1200
 agtccatcct tcaactgaaga tgggatgttt attttttaat taattaatta atttattttt 1260
 tgagacagag tctcgctctg ttgccaggc tggagtacag tgggtcagtg gtgcgatctc 1320
 ggctcactgc aacctctgcc tcccgggttc aagtgtattc cctgcctcag cctcctgagt 1380
 agctgggatt acaggagcct gccaccatgc ccggctaatt ttttgtattt ttaatataga 1440
 cagggtttca ccatattagc cagatctcga tctcctgacc ttgtgatccg cctgcgtcgg 1500
 cctcccaaag tgctgggatt acaggcatga gccactgtgc ccagccagat gtttactatt 1560
 atgtgtctgg atacattggc aaacaagatc gctgttactc tttttttttt tttttttttt 1620
 tttgagacgg agtctcgaac tgttgcccag gctggaaggc tggagcgcag tggcgcgatc 1680
 tcaactcgga gatcactgca acctctgcct cccgggctca agcgattctc ctgcctcagc 1740
 ctctctagta gctgggacta cagggtgcacg ccaccacgcc tggctcattt ttgtactttt 1800
 agtagagacg gcgtttcacc atgttgggtc ggatgggtct tatctcctga cctcgtgatc 1860
 caccgcctc agcctcccag agtgctggga ttacagggtg gatttgccgt gcccggtatc 1920
 cctagctgtt actctttacc agatccagta taggcccctg ggactggagt ccaagcctag 1980
 aatgggttctc ctgggtccact ttctggcctc tttctcctgg tgtgacctca aaccccggtt 2040
 ccctattatc cctctatca gaaaaatgga gacgatgcac actgcccacc tggcacacaa 2100
 gggaggagga aatcgacact tctgaagatc tatggttctg caccacactc cacagcctgc 2160
 accactacct tggttcagcc agcataggag tccctagctg cttctctctt gtcccagggt 2220
 tctgtttcca gaaaggagag gactgagacc cagcagttat ccttccagggt tctgggtgta 2280
 taacacagct tgttttactt ctaaaaattt agtgcacgct gtgtacctag 2330

<210> 4
 <211> 266
 <212> DNA
 <213> Homo sapiens

<400> 4
 cagagtgaga ccttgtctca aaaacaaaaa tacctagaaa atcaatctac tctgtctttt 60
 aatgtgaaat gttcttatga tagctatctt tcttagtttc ctttttttct gaagcactaa 120
 acacaacctg taggtcttat ctctggggtc tgggaaacag aaccttaatg ttacagggtac 180
 aaaagcaaac agagtgatta gttccccatt ttctggtagt gaacaactga cattttttca 240
 atcttatgta aaatgtgaat aaaaat 266

<210> 5
 <211> 1483
 <212> DNA
 <213> Homo sapiens

<400> 5
 atgagtagtt ttcctagctg attttttttt ttttaataga ttttgaagtc agatggataa 60
 aaatgccatg tggacaagtg cttacttttg cattatgaaa attttataat ttaacccaaa 120
 caaaagtga atataattc ttgagtttca tgcctttga ggtgcctttt taaaaataat 180
 caaaatggtg ttgggagacc ccatccaatt taatcgggtg ttatttaatt atactactat 240
 aattgttgta tttgcagggt tgactgttct cagggaacgc tgaagggttca taacagtagt 300
 gatttgtaat tgtgaggctt gagggtggaa ttgaattact tcattaaaga gtaaccagtt 360
 actgagaatc catgtttcat tgagactggt gtttttctac tttatccttt tttttttttt 420
 tttgagacgg agtcttgctc tccagcctgg caatcagagt gagacctgt ctcaaaaaca 480
 aaaataccta gaaaatcaat ctactctgtc ttttaatgtg aaatgttctt atgatatgta 540
 tctttcttag tttccttttt ttctgaagca ctaaacacaa cctgtaggtc ttatctctgg 600
 ggtctgggaa acagaacctt aatgttacag gtacaaaagc aaacagagtg attagttccc 660
 cattttctgg tagtgaacaa ctgacatttt ttcaatctta tgtaaaatgt gaataaaaaat 720
 aatttttagaa aagttatcta tttttatgtt ctgtaacaca aatagtttaag aaaatgaata 780
 cttgttatgt aaatgaagct tcacacgagg acctccggca taactttgat catgttgat 840
 ctcttaagca ttttatatag gaattctggt gtggctacct aaagcagaag ggaaggcaat 900
 caactaaaaa tccctttcta ttgagttatt ttcatcatgt aaattacttc aggtttttct 960
 tgttagggct tcttgaatgc acatttgta ttatctgtat accaaagggt agcttttttt 1020
 tttaatccat gtgattttta tgagttgaaa gctataaggt tttttaatta aaaatttctt 1080
 tctaaatggc aaatttggtg gacagcattt aagacactca gtttctattc aaaagcaata 1140
 agaaaacaag ttatgttcac gggttatgtc ataataaggt ccctgggact ttattaaact 1200
 gtcatctatc atactgctat gaaatctggt cctagttttt aacagatcct ctgaccagtg 1260
 ggttactccg ggacttcacg tgccccgttt cccaacacct cgtcctgtct cggctacact 1320
 tgaagcttat ggtctttcca atgcgacagc ggggttgga acggcatata ttccggcggt 1380
 gagggccctt gcccgaagag ggggacgcgc ccctgcgga ctttttccgg tggggcaccc 1440
 tttggggcat acaactcccc gaggttgta taagtatact ata 1483

<210> 6
 <211> 345
 <212> DNA
 <213> Homo sapiens

<400> 6
 tatctttcca tgctgtctgg ccttggtcag tgcattocca gccatctta actcattttt 60
 ctctccttga acctgagaga agatgttgca gtcctagtcc ttttctccc agggagcatt 120
 tctgcccttt gcaaattcat ggatgctaaa caaaatgtgg agaaaacata ttgccctgct 180
 ctatctggca gcttccaaga ttcaatgata tattgggaaa ggagtaactc acttccccct 240
 ccagcaacat gtaagcccta gactcctgcc aggccaaaaa tatccccgat taaacaaatt 300
 atgtagcaga aagtctttga attaataaac caaatttgaa taatt 345

<210> 7
 <211> 491
 <212> DNA
 <213> Homo sapiens

<400> 7
 acaaccatac cagaaggggt tccaccattg gccatactct ataatacaga ccaatgctct 60
 tcacctacaa tccggacact ggtatagcac ttgtcagggg tccagactct ccctctttta 120
 tagtaggttt cctcttgagc ctttattatc tttccatgct gtctggcctt ggtcagtgtc 180
 atcccagccc atcttaactc atttttctct ccttgaacct gagagaagat gttgcagtcc 240
 tagtcccttt cctcccagga ggcattttct cccttgcaa attcatggat gctaaacaaa 300
 atgtggagaa aacatattgc cctgctctat ctggcagctt ccaagattca atgatataat 360
 gggaaaggag taactcactt ccccttccag caacatgtaa gccctagact cctgccaggc 420

4

caaaaatattc cccgattaaa caaattatgt agcagaaagt ctttgaatta ataaacccaaa 480
 tttgaataat t 491

<210> 8
 <211> 91
 <212> DNA
 <213> Homo sapiens

<400> 8
 cccacgcgtc cgaaagatag gccaaagctgg taccaagcat ggatggattt gtcaaagatc 60
 aggccacctc atctcttcca ttggccacca g 91

<210> 9
 <211> 890
 <212> DNA
 <213> Homo sapiens

<400> 9
 cgtaatttgg agttctcctt cgaatttgac caactcaggt agagttaatc aaatctgatg 60
 gaagaaaaacc aaaataacaa aacaaatattg attactgagg acttttaatg gtaaggagaa 120
 gttaaagacca gttacttgtc aatcttaact ttagtcact aaggggaatt ttcaagacaa 180
 aactctaatt gagctactta cctaggaatg aggtcacgc tgaacactgc tgtctaccat 240
 ctatgaagca ggaaaaaact caaactcact ttctctgttg gaagggagca gaaactccag 300
 aaaggacttg ctggccctcc atcatcatgg aaacaggaaa actaatcttc cttgttggaa 360
 gtgagtaaaa ctccaaaaa ggaggagtgtg tacagcaaaa tgaaccttag acctcaacca 420
 aatttgggga gagcaatgat tctctgaagg gacctcccag acctcagcaa attgtattat 480
 tggtttgagc aataaagata ggccaagctg gtaccaagca tggatggatt tgtcaaagat 540
 caggccacct catctcttcc attggccacc agtttataaa ccaaagagta tctgagacag 600
 gtctcaatca attcagaagt ttattttgcc aagggttaaag acatgcctgg aagaaaataa 660
 catggacca caggaacagt ctgtggtctg agtctttctc caaagataaa tttcaaggct 720
 tcaatattta aagggaaaat gtaggctgga ggggagaggg gtagagtata gtaatcccca 780
 tgttgtaaga gaaaaggagc gggtagggga atagtcaatt atgtattcat ctcatgctca 840
 ataaattggc actttacata agaaaaaaa aaaaaaaaaa aaatgcggcc 890

<210> 10
 <211> 386
 <212> DNA
 <213> Homo sapiens

<400> 10
 gactgaaatt agaaatgaaa aaaaagggtta ttcattgatt gaattagcag tgtctggcac 60
 acagtaatgt aatatataaa gtaagttggg cttggatgcc atctaaaagg gcttagttat 120
 tgaagcagtt ttatttctga agtgatacta aagaataggc atgcgtgtcc aggctagtta 180
 cagttaaatt taggaataag gcacagtaat aaagatacta tctttagact tgaaaattat 240
 aaatccttag ttgttattcc ttatcagttt ataaagttaa caatgaatgt acagactaca 300
 agctatcaaa tgtactgtag atgaaaaggg caataaacac tagtcagagt tagaggtaga 360
 atgatataat ggaaaggtag gaattg 386

<210> 11
 <211> 458
 <212> DNA
 <213> Homo sapiens

<400> 11
 gaaaaaagggt ttatatctca caattgagtg ttccgttgta gaatgtctct tatatatctt 60
 tagagtgtta ctatgtcttg ttcaagtagc acaggggccc ggaaatacaa tttgaaggga 120
 gaggctaatac tttaaaggca gtcattgatt gtatttttaa ttacaattta caaccccatg 180

5

```

gtaatgaaca cataggctaa acaaatataa actcaattaa aataacatgc aatgatactt 240
tacaaaaaatg ctgctgagaa gaacatcgag ttacatcat gctgaatatac taaaaatagc 300
tagatgacta tttaaccttc tatatatatg tatagatata gcgttataat tttcccacta 360
gaatttaatt ttatattata gaccctttca gtgccttcag tgaccctatg agtgtctttt 420
taagattgcc tttggaccct ggtgtcggtt tgggactg          458

```

<210> 12

<211> 490

<212> DNA

<213> Homo sapiens

<400> 12

```

gaaaaaagggt ttatatctca caattgagtg ttccgttgta gaatgtctct tatatatctt 60
tagagtggtta ctatgtcttg ttcaagtagc acaggggccg ggaaatatac tttgaaggga 120
gaggctaattc tttaaaggca gtcattgatt gtattttaaa ttacaattta caaccccatg 180
gtaatgaaca cataggctaa acaaatataa actcaattaa aataacatgc aatgatactt 240
tacaaaaaatg ctgctgagaa gaacatcgag ttacatcat gctgaatatac taaaaatagc 300
tagatgacta tttaaccttc tatatatatg tatagatata gcgttataat tttcccacta 360
gaatttaatt ttatattata gaccctttca gtgccttcag tgaccctatg agtgtctttt 420
taagattgcc tttggaccct ggtgtcggtt tgggactggc actaagtgc atcctaaaaat 480
tttcataaaa          490

```

<210> 13

<211> 64

<212> DNA

<213> Homo sapiens

<400> 13

```

agaaatgtaa atgctatatt agaaaatatt tactaagtcc aagagaaagc aataatagag 60
gcac          64

```

<210> 14

<211> 921

<212> DNA

<213> Homo sapiens

<400> 14

```

atcacttttg ttccgggctat tgtctctgat gactttctat tgctccctat ttaccctctg 60
ctgtcttnag aaaaggaact agagaagtag aggggtcgatt gtagggctgt gtctagggca 120
tattccctgg aaataaaata gggtcttgga gctgtactga gagcagcttt cgaccatgct 180
aaattcctat tagtagtttt ttttaaata accaatttgc tattaatatg tattctttgg 240
tgaaactgtc caaatatttt gaccatcttt tattttatta ttattattga tatagctgga 300
ttcaggagtg ccaactttat catTTTTtagt gtgtctctta tcttttttgg cctctatta 360
ttgctttctc ttggacttag taaatatttt ctaatatagc atttacattt ctttgatgat 420
ttttttccac atatattttt acttgatcc ttgtagtgc ctctaggaac ttacaatgta 480
ctatttttat tattattttt ttttaagaga gagtctcact ctgtcacttg ggctagaatg 540
cagtgggtgt accatggctc accagaccat caacatcccg gggtcaagca attctcctgt 600
gtagttggga ctacaggtgc gtgccacaat gcctggttga tttttgtgtt ttttagtaga 660
gacaggggtg caccatgttg gccaggtcgt tctcggtc ctggcttcaa gtgatctgtc 720
tgcttggcc tctccaagtg ctgggattat aggtttgagc cactgcaccc agcaaaaaac 780
caacttttta aagcagaact tctgaagaaa gataaatatg tattaatgtt gactatgtga 840
cagatgccat gctgtgtatt acatgcttct gttcatttca tactccctag taatactgat 900
aatgtttttg gatagtcatg g          921

```

<210> 15

<211> 270

<212> DNA

<213> Homo sapiens

<400> 15

```

tgaacatcgt ctttcttagg tgacttcctc cacatagtta tttgtgaatt gtaatattgt 60
gtggcaatat tatccaaaaa gtatttggtt ttattgtatt ttgcaatccc taggacatat 120
taaatagctc aagtgatggg atgttcttct taaaattcga tctcatagtc taaaatgggtg 180
acttgacatc gagttagcct cattaagtcg tggacgaatg gcaaattgcta taatttctctg 240
tctcacttct ggatatcaca gtgtcatctt

```

<210> 16

<211> 651

<212> DNA

<213> Homo sapiens

<400> 16

```

cgatggagct tccgagggga aacggctggg tatcttataa gtcttgaggg ccttcactcc 60
cccaaccatg tcttttgaac attttttatt ttcttttag agacaggatg tctctatgct 120
gcttaggctg gagtatagtg gtgggatcat agctcactgc aagctcgacg tcttggtatc 180
aagtgaactt actgccttga cctcccaaat agctgggact acaggcgtgc accaccatgc 240
tcggctaatt tttacaatgt ttatgcagat ggggtcttgc tctgttgctc aggcttgtct 300
caaatcctg gctcagatg atcctcctgc ttttgggtcc caaagtgtg ggattgcaga 360
tgtggctcac cagcccccgc ctgaacatcg tctttcttag gtgacttcct ccacatagtt 420
atttgtgaat tgtaatatgt tgtggcaata ttatccaaaa agtatttgtt tttattgtat 480
tttgaatcc ctaggacata ttaaatagct caagtgtggt gatgttcttc ttaaaattcg 540
atctcatagt ctaaaatggt gacttgcaca tgagttagcc tcattaagtc ctggacgaat 600
ggcaaattgt ataatttcct gtctcacttc tggatatcac agtgtcatct t 651

```

<210> 17

<211> 702

<212> DNA

<213> Homo sapiens

<400> 17

```

gcaaaatagc ccaggctgtg tgctgaaggc atcccatctc ttctctgtcc cctctcggct 60
ggatcggggg ggggcagcgg gtggatgagt gtgtctgtcc tgccagtcca gcctccaact 120
ggtctgctgt ggggcaggag cccacctggc tctctgcag agctgcatgg cctgccttgc 180
ctcaccctg acaacagaga ctttggttct ccactgcag acgcatttgt cttgtttctt 240
attcgggtcca gaactcgggt gggaagaagg gtgatgtgat ttgggtcccc tcaagacctt 300
gacaacagat agtttttaat atcacatttt aaagccgcca actttctctc ctccactttg 360
gtatttccct gatttttaaa cagaatgtcg gctctggagg cagaaagctt gggcttggtat 420
ctgggcccgt ccacgcagta gctatgtggc cagggaatac atagcttccg gatctccgat 480
ccctacagta aagtacagat aaaaataact ttcatgtatg tgtttgaaat cgaatgagat 540
agttaatgaa tgagtaagt ctctgcaaac tccagagcgg ggtgcgcgtt ctgatctgtt 600
tcatagaatc tgacacgtac cctttccac cccagcgtct ctgaattggg atgcatctga 660
cagcaagtgt ggcacccggg ctgcagttgc cgttgtctgc tc 702

```

<210> 18

<211> 1760

<212> DNA

<213> Homo sapiens

<400> 18

```

gtccaccgtg aggatgggtc caccocatag ctttgttgat gctgtggctg tagcaggccc 60
agtggctccg ctgagcctcc cttagcaccag tgcccagtga agtgtggggg gctggacata 120
gggtggcctc gttcctggga ctccccaaa gctctgggtt cccccctgct cccaggcttc 180
aggtggacga gttagtgaac acccccactc cagacctccc tcccctagcc acccccacag 240
ttataaaaac cttcattctc atatggaacc ccctttcctg aaatccgtag agtgactgac 300
tgctttcttg agtgaatctg gactgggcca catgatgac gctgtacaga gcagctgagc 360

```

```

tcctctgtct cagcctccct gagttcacag caggctctgg gcacatctc cgtgtcatcc 420
taaggccacg ggcgggggtc ccaccaaaca ggagagcagc tctcccgaga tgaagccttc 480
tgatagccct agaaccaaga ggaaccgtgt ggggttgggt ggggtgttta ctgtgactc 540
ctgatgttcc ctcccagtgaggacacca cctgggacac tgtggcccct ggccctccct 600
ccctccctc ggtggcagag agaacttcct ggtgggtgac agcccatggt ccacccttgc 660
caggtggatt caggcaaaat acgccaggct gtgtgctgaa ggcacccat ctcttcctcg 720
tcccctctcg gctggatcgg ggtggggcag cgggtggtga gtgtgtctgt cctgccagtt 780
cagcctccaa ctggtctgct gtggggcagg agcccacctg gctctcctgc agagtgcac 840
ggcctgcctt gcctcacccg tgacaacaga gactttggtt ctccatctgc agacgcattt 900
gtcttgtttc ttattcggtc cagaactcgg gtgggaagaa ggggtgatgtg atttgggtcc 960
cctcaagacc ttgacaacag atagttttta atatcacatt ttaaagccgc caactttctc 1020
tcctccactt tgggtatttcc ctgattttta aacagaatgt cggctctgga ggcagaaagc 1080
ttgggcttgg atctggggcc tgccacgcag tagctatgtg gccagggaat acatagcttc 1140
cggatctccg atccctacag taaagtacag ataaaaatac ttttcattga tgtgtttgaa 1200
atcgaatgag atagttaatg aatgagtgaag tgctctgcaa actccagagc ggggtgcgcg 1260
ttctgatctg tttcatagaa tctgacacgt accctttccc accccagcgt ctctgaattg 1320
ggatgcatct gacagcaagt gtggcatccg ggctgcagtt gccgttgtct gctcacatgt 1380
gaattaaaaa aacaatctca gcatatgaaa tctctaactc cgtatgaact tgggtggtat 1440
tcctgggtgcc gtgtggataa ctgcagcctc aacaccccag tccacaaacc acgtacggac 1500
caactgagga aggagtaggg gttcttgttg ttgcagaaaa cctccctca acttgcctca 1560
gaagatacca gcattcatac ttggagtggg catcagcagc ttggaagaca cagcagtggg 1620
cccctcttag caggagtgc ccatctcatg ggctcccagc gacgaaccag aggggtgatg 1680
ctgcgtaggg gctcacggac attggcactc taagtcagag atgctcaggt gaagggggct 1740
ctgatggtga ggacattcag                                     1760

```

<210> 19
 <211> 284
 <212> DNA
 <213> Homo sapiens

```

<400> 19
gataattttcc ttcagagacc ctttacattt gagagatcct ttagactttt gtaaaactta 60
ataatttttaa ttaataacta gcgttaattg aattcttcct ctgtgcaagg ccacgttcta 120
agtgccttcc tcacagcaat gctgtgaagt acttatcctc cttgttcttc tgaggaaaca 180
aggctgacag gcctgagatc acagagccag taaatggtag agtcaggaat tgaacctgag 240
aattctgact ccagactttc ctgcttttagc caccgtgcag tact 284

```

<210> 20
 <211> 1150
 <212> DNA
 <213> Homo sapiens

```

<400> 20
gataattttcc ttcagagacc ctttacattt gagagatcct ttagactttt gtaaaactta 60
ataatttttaa ttaataacta gcgttaattg aattcttcct ctgtgcaagg ccacgttcta 120
agtgccttcc tcacagcaat gctgtgaagt acttatcctc cttgttcttc tgaggaaaca 180
aggctgacag gcctgagatc acagagccag taaatggtag agtcaggaat tgaacctgag 240
aattctagac tccagacttt cctgcttttag ccaccgtgca gtactgccta ttggtcagtg 300
taccctgaga tactcagttc attttagttc ctctaaagtt ttgttattaa aaagtactg 360
taaatgcatt gtgtccagag cattatagca tacttttaaa aattattcac ttcttaagaa 420
ttctactcat cccaccctca tcttttgaaa attaacactt tacctacatg acttaaaatc 480
atctgaagac ttttaataag ttgctgagtt tcatgtttca aaacctgtta tctactactg 540
gagcaattaa aattaaccat acaacaggta acaggtttaa gtgactttgc cttggtttta 600
actaagcaca ggttttaagt ttgtaagcgt ggataggttg ggagcaagct ctctagtggg 660
aatggatttt aaacctaaat agtaagtga aaccatgcag aggcgtgctt gtcgctgtga 720
gactgtgctg tatgtgtcta gactggtgga gcagtacaga gaacagagct ggatgactat 780
ggccaatttg gagaaagagc tccaggagat ggaggcacgg tacgagaagg agtttgagga 840
tggatcggat gaaaatgaaa tggaagaaca tgaactcaa gatgaggagg atggtaatat 900
tatttttatt ttatttatct tttttgtttt ttaagtgaag ctggaaatct ccttgcctat 960

```

8

```

ttgacatctc ccaatTTTTa aatgtggcaa ataattaaaa ataatgttgt atggggccaaa 1020
ggtagtgggc tgagctagtc taattcaagt aatttgatta acaaattctt ttctgaccat 1080
gtcctaaaca gtgtgtactt ctagctgcat aatatgacaa atggacatgt ttaccagtgt 1140
gactatTTTT                                     1150

```

```

<210> 21
<211> 226
<212> DNA
<213> Homo sapiens

```

```

<400> 21
aaaaataaaa aaattcaatg aatcctgtaa atcctttcat tataaaataa atttggattt 60
gatatacaat tatggcctct gagtagcctt tgaatcatct ttagattcta aacttaattc 120
tgaaaatatg ttttaccata gtataaaata gtttttatgt ttatattaga aaaatgatgt 180
ttaaatTTTat ttctaagaat tacttttaggc caggtgcaat ggttca                226

```

```

<210> 22
<211> 270
<212> DNA
<213> Homo sapiens

```

```

<400> 22
gcgTggettc gattccggcg cctgcgtgtc accagcccag ggtggccgtg gaagctggac 60
ccgagccgca ggccccccag gctgggcctg ggaggaaagc ggtttgaaaa agatcggaac 120
tgaggaaact tcttagagcg ggggactccc tgctcctaca gccttaacca atgcccagcg 180
cttggaaggt ggaggactcg gggattcggg agcgtttcag gcctggggaa atggaagggt 240
cggggaccta ggtgaaaggT tatttgccag                270

```

```

<210> 23
<211> 245
<212> DNA
<213> Homo sapiens

```

```

<400> 23
ggcacttgga ttgtctccat tctctgcacc caagctgtca gggccctcac cagaatgttt 60
acctaacacc ttctctctag tctggagtct ttgtagatgg aaaacttgat gtataaccct 120
ttgacttgat ttccaagaag caacagagtt aaaactgtta tttctaggtg agtggcttca 180
tgcaggtgtg gtcaggtatt tttcctgaca gaggtgtctg ttcttggtga ttgctttttc 240
ttttt                                           245

```

```

<210> 24
<211> 460
<212> DNA
<213> Homo sapiens

```

```

<400> 24
atTTTTggtt ttaaattcca tacattctag tatttttgag acttttTcact gcaaatttta 60
acatgcaaaa tgtacggcct ggTttccata agcataaaata gtataaatgc caacaataag 120
aatgtcttct aagcagctaa atcttgtaag tttagtTgga attgagacca gctattTggg 180
taagcgaatt agagtcttag tattgtaagt gggTatgttt atgtggcaca gggTtgccaa 240
ctgcctgagt ctattcgtga gtcagaacga ctttgctgat gtgtTgggcc aagccagccc 300
tggtTggcag cctggTgcag ccgtaaaaatt cagcctTaca aacagtctcc cgccattccc 360
gcaccatggg actTtagtgt tgtgtgtaac aacagtataa cctgctgtta gccattatc 420
aactgactgc tatgctaaac caaaattata ataataatgc                460

```

```

<210> 25

```


<211> 257
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (93)..(192)
 <223> a, c, g or t

<400> 25
 gtataatgat actaatcgta aaaacaaaaa aaatctacta agtacttacc atttggttaga 60
 cactggggttg agagttttat atgcattgtc tgnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnaacctcag aggccaagct cttaggcact gtgatatact actggcctttg 240
 ctcagtaaatt ggacctt 257

<210> 26
 <211> 221
 <212> DNA
 <213> Homo sapiens

<400> 26
 ctcgagaccc cacccttcc tggattcatc agtgggctcg gaagagcgtg ggaaaggcgc 60
 taccttggtt ccacaccacc tgagtcagct ggggctactc ccagctctcc ggggtgggaa 120
 actggtgcct ctggcaatgg cctgctgagt atttaacccc aggggcagca gattccttgt 180
 ggggtgtttt ctacaaatta aacagggaagg ttttttgcag g 221

<210> 27
 <211> 347
 <212> DNA
 <213> Homo sapiens

<400> 27
 tgcttgcttg gctctggctg gggttcgtta gggctggggt tcttgaaggg ccctgtctaa 60
 gaagggagta ggaatccagt tatatgagtt cagctcatc aggaacctgg catatttgat 120
 tgagagatat gtccagtgat gccctgttg aagctgctca tgaacagggc ttggtccttg 180
 acacttggtg ggcaagtaat ttacagggga aatgacaatg ttaatcctgg cccctggggt 240
 gctggcagtg tggtaagga gacccaacac acacagggat gggacccaac acaagctaag 300
 gaagggcca cccccagccc tgatgtctgc tggaacaaag agaaatg 347

<210> 28
 <211> 338
 <212> DNA
 <213> Homo sapiens

<400> 28
 tttttaaat gtgaactata acacttagga tattgcatgg atcatcaaaa aagataaatc 60
 atctctttta aattctgtgt tattttaaaa aacaaataat agatacagat gtctgagtat 120
 tttaagacat tttggggtt ctagtaatta ttagtgccat taaccacaaa gacaaaggaa 180
 ggggtctgtc ctttttaaat acagtaatct cactgtagag ttcaagccat gaggtcacaa 240
 gtatcttaat attgtacnaa aaccttttct ttttcattct agcctcttaa cccctaagca 300
 aaacaaatga aaaaaatgta ctaaaaact taatgttt 338

<210> 29
 <211> 622
 <212> DNA
 <213> Homo sapiens

<400> 29
gcctgaagct gctctctagg aaaatgtggc attctctgct tgggggagggc tgggggtgggg 60
gtaagagaga gggaagatgc cctcagctcc caccaaggag cataaataaa aagagaattg 120
acccccagc acccttcaat agcccaccag agttgccacc aaacagtgtg aaaacgtgtg 180
gttttgacta ttctgatgaa aataatggat gttctgtgga gatttgtaga gcacacacac 240
atatgatttc taaatcaa atcagttgcaa ctgttcccat cagaaagacc catcaagccc 300
ataaaagaga tcccttcata caaagatctc ttgcatccc aatttccacc cattctacat 360
gcattttcaa acccatttcc tgatttctact gtcattagct agaaagcagg gggctattag 420
cctggattgt aaggcatcca tttctccttt ttttgtttca ttagccatgt aggaagatat 480
tttctcttta tggttgatgg catctgtttt taaaaatgga taaactcttc aaaacatagt 540
ttctgattct ggtagcact agatgagcag ctgtaaaata ataataatag tttgaggggt 600
tgagaagagc tttctttatt tt 622

<210> 30
<211> 518
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (260)
<223> a, c, g or t

<220>
<221> unsure
<222> (262)
<223> a, c, g or t

<220>
<221> unsure
<222> (333)
<223> a, c, g or t

<220>
<221> unsure
<222> (337)
<223> a, c, g or t

<220>
<221> unsure
<222> (343)
<223> a, c, g or t

<220>
<221> unsure
<222> (354)
<223> a, c, g or t

<220>
<221> unsure
<222> (371)
<223> a, c, g or t

<220>
<221> unsure
<222> (376)
<223> a, c, g or t

<220>

<221> unsure
 <222> (380)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (470)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (304)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (310)
 <223> a, c, g or t

<400> 30
 cagatcccca aattcctctc caggatgggt gcacgtggcc cctcaggaac cggggaagtg 60
 cacgtgtggg tggagagggt tgaggaaaag agccagcttc cggacacggg tgcagggtct 120
 ccagcagctg agctcccgga gtgtcaagtt gccggagggt ctgtgcctga gcaagcagag 180
 aaggaaactt aagcctctaa tgaaaaggcc tcctgttctc ttgcaggaga agccccaga 240
 gggtaatggg gcagtggccn antggcctgt ggtgaccca aggaggggga ggggccaggg 300
 ccantgggn cctcagaata ttgttcctgt gtnttonttc gangcgggtc tggncctgct 360
 ccgcagcctg ntgggntcan gactgaacag tctcctctca gcctcatggg cggttgtctc 420
 tgggcacagg ctactcttaa cctcgctcc ttaacccac acagggcagn ctctgctgc 480
 taaaaatatt tctggggaca cggctctaaa aatgacct 518

<210> 31
 <211> 556
 <212> DNA
 <213> Homo sapiens

<400> 31
 cagatcccca aattcctctc caggatgggt gcacgtggcc cctcaggaac cggggaagtg 60
 cacgtgtggg tggagagggt tgaggaaaag agccagcttc cggacacggg tgcagggtct 120
 ccagcagctg agctcccgga gtgtcaagtt gccggagggt tctgtgcctg agcaagcaga 180
 gaaggaaact taagcctcta atgaaaaggc tcctgttctc cttgcaggag aagccccag 240
 agggtaatgg ggcagtggcc tagtggcctg tggtagcccc aaggaggggg aggggccagg 300
 gccatctggg tctcagaat attgttcctg tgtcttcttt cgacgcgggt ctggccctgc 360
 tccgcagcct ggtgggctca ggactgaaca gtctcctctc agcctcatgg gcggttgtct 420
 ctgggcacag gctactctta acctccctc cttaaccca cacagggcac gccctcctgc 480
 tgctacaaat atttctgggg acacggctct aaaaatgacc ctgccttcca ttcactggac 540
 agtgaacaca agaattg 556

<210> 32
 <211> 330
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (151)..(176)
 <223> a, c, g or t

<220>
 <221> unsure

12

<222> (247)..(273)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (311)
 <223> a, c, g or t

<400> 32
 tctgtgcttt gtgtaacttt ttgctaaatt cctgtctttg tcttcttggg acagtcttct 60
 acttggtaca ggatcttctt atcttttggg ttttatatta gttttaatat aaaattaata 120
 tagttttata ttatatagcc cactgacatg nnnnnnnnnn nnnnnnnnnn nnnnnntgac 180
 ttggccagag ccttcagttt cttatctctg gtaagaggta atgtgtctct ccctagggca 240
 aggctgnnnn nnnnnnnnnn nnnnnnnnnn nngatgtgt gagagaagca gggagagtaa 300
 gaatcaagac naaactgcag tcttttatac 330

<210> 33
 <211> 431
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (420)
 <223> a, c, g or t

<400> 33
 aagacagcta agtaagtggg ggtaggaaga aagactggac aagggtttga tggactggct 60
 atgaaagatg aggaagagag aagtcacagt tgggtaagag gaagttttta aggaccacca 120
 agaaaatggt gacactctta ttagataacc tagaaattag acaaggatga gatgttatct 180
 ggatattcaa atgaaaatac cctctattca gctatagtcg ggctactggg gttttaaggg 240
 agaatttcag atttgtggaa ctcagagagt cttttgcatt tcaaagaagt gataattgag 300
 aagctgtgtg acaactaagg ttgtactaga agaagcctag acgtgagagc aggaagaatt 360
 catggacagt gctaagttag gacatatatg ttacacagat gacaccagtc tggatgttgn 420
 agcccagaca c 431

<210> 34
 <211> 275
 <212> DNA
 <213> Homo sapiens

<400> 34
 atttgattaa ttttgctttt gtagtttgtc ataaaaccac agtcactggt tcattacaat 60
 taaagataat tgggtacgct actcctgagg gaaaccagca ttcaaaatgc atcccccca 120
 tagtttttat tatttgtgag agaatgtctc attaataatt tcagagcatt ttggatttca 180
 aaatatttgc cttagacctt cttgcctcct cttctcttgt agagccatat gggtcctttg 240
 tactcagaaa attgaaaatg agccagggtg cagtg 275

<210> 35
 <211> 497
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (486)
 <223> a, c, g or t

```

<400> 35
agtgatttca ttatctccaa tgtgtatggc ttgatagaaa tagattccat tatgtagcac 60
cttaaatacca gataaaacat aaggaatttc tattccatgt ttgtatgac aatgttaata 120
atctaagaaa atctaaaaag aagctacttc ctctattaca gtatgaaata aatagtctga 180
atgatttgtt ttgggggggtg gaatggaaag gtataagact gaggagggtg cctgtgggaa 240
cagtgatagg aatcctttct taagggttgg gttttacata cgtcttttaa aatagatgat 300
atcattaata aattatctgt gggcatcatg aaaaaagtgt ataacgtaca actttatgag 360
cttgacagtt ggtgaaaact tttctgttta aaattttatt tggccctccc caaaagaaat 420
ggttatttat gagtattagg atagttccag cagtaatgcc tcaaaagaac caggagggtat 480
agtgtngtct aaaatgt                                     497

```

```

<210> 36
<211> 1796
<212> DNA
<213> Homo sapiens

```

```

<400> 36
tgcattctagt ccaccacctg tttttgtaaa gttatcagaa cacagtcattg cccattcatt 60
tacaatttgt gtatggcttc tttccctgca acagcagagt tgagtgttg aacagaaacc 120
tatggcctgc agagttttaa atatctaccc tttggccttt tataaaaaaa gtttactgat 180
tcctgggtgag tatattaaaa agttaggaaa acctaaatct tccagagtgg agaattagaa 240
agtaagacgt gttgtatata agacagacag tttgtgtgtg cgtttattta taaatatatt 300
attctgaaat aatggtgtcg acatatgttg caggtcttaa aaattggtca atatatagt 360
ttaatcaaaa aatggcaaat tgtaaaatgt agacagaatg tgattgtgta tttgtgtgat 420
acaccaacag aaaagggtgc taggaaacct gtggaccaac ataactaagt tggtcttttt 480
gatgggtggt tcatggattt ttaaaaatct tcttggtttt ctgtagattc tgactttcct 540
gtaatgagta tgaataagta tgtatttctt gagaaatgtg aaaataactt tatcttccca 600
gatttctcat aattgaaaat gttggaataa atggtcctgg gacagatctt tccattgaga 660
agggcggaag ggaaaccctg gggattcagc tgggtttctg ttgcatttct ggtaacacac 720
agttgtgaaa agccagtgtt ggccattccc caggacagtc tggggtagag gaggtcagga 780
tttaactact tgagggtccg gggaacagat gtggccacag tccttcctga ctactgttt 840
tccttccac agtcccctgc ttctcttcac tgatgcacat agatgcctga ccagaggaga 900
gatttagttt tctccaagg attatctgtt atgttgcagt tctgaaattc ccataacgtt 960
taggctagaa cacaagtgtt ttcattatct ccaatgtgta tggcttgata gaaatagatt 1020
ccattatgta gcaccttaa tccagataaa acataaggaa tttctattcc atgtttgtat 1080
gatcaatgtt aataatctaa gaaaatctaa aaagaagcta ctctctctat tacagtatga 1140
aataaatatg ctgaatgatt tgttttgggg ggtggaatgg aaagggtataa gactgaggag 1200
ggtgcctgtg ggaacagtga taggaatcct ttcttaaggg ttgggtttta catacgtctt 1260
ttaaaataga tgatatcatt aataaattat ctgtgggcat catgaaaaaa gtgtataacg 1320
tacaacttta tgagcttgac agttggtgaa aacttttctg tttaaaattt tatttggccc 1380
tccccaaaag aaatgtttat ttatgagtat taggatagtt ccagcagtaa tgcctcaaaa 1440
gaaccaggag gtatagtgtt gtctaaaatg tggactcagg agccagactg cctggctgtg 1500
caactagcct tgtcacttcc tagatatgtg gcaagttaat taacttctca gtgttcttat 1560
ctgtagaatg gggataatcc taatatatcat ctcagggtta tattacaaat ttaaaaagtt 1620
aattttgtaa aggacttaga atgatatctg gcaaataaaa gtgttcataa aagtaaacc 1680
tataaaaagt tttactcatt aaatacaata atctgaaacc attagtaatt taaacatttg 1740
tggctgactt ggtaatatatt atgaaaataa atactgtatt tataatcttt gacctt 1796

```

```

<210> 37
<211> 83
<212> DNA
<213> Homo sapiens

```

```

<400> 37
gttgggatct gaaagaggaa tctgtggata ctgaggaaag gtagccagaa aggttcaaag 60
taacgccaag aaaaaatggt gtc                                     83

```

```

<210> 38

```

<211> 773
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (295)..(592)
 <223> a, c, g or t

<400> 38
 ggacaacaac caagggattt ggcccaagaa gaagaaatat aggcaagagg aaaaaaaaaa 60
 aaaagagaga gagttataga atagagtaac agatttggaa atgcatcaat agttgaaacc 120
 tggagagcag ataaaattac ccaagtagag aatgtagagt aaaaagaaag gaaaggatg 180
 gacagaaccc tgacaaaaca ccaggattac agttgggacg tgaaagagga atctgtggat 240
 actgagggaaa ggtagccaga aaggttcaaa gtaacgcca gaaaaaatgg tgtcnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nntgggtgtca 600
 taaaaccaag gagccacata aagttttaag aaggaaaaaa tgtccaacca tgtcatatgc 660
 ttccaaaagg ttaaataaga tcagaagtgg aaattattat ttgaacttaa caacatagaa 720
 tccttaagga cagttgtgga atttcactgg aatgcgagtg acaattgaca ttt 773

<210> 39
 <211> 326
 <212> DNA
 <213> Homo sapiens

<400> 39
 gaagtgaatt tacagaatct gagcatggat tagttgtata acaggggtgg tgggtcttga 60
 gggcaggtag caaagcaaag aacgacttga aggtttgaaa ttgaaattct gaatggacct 120
 ggatagcatt taatgtgata ggagaaacta tgaatgaaat atgaatatct ttgttctaca 180
 gggagttgag tgggggggat gaagatagtt aattttgaat atcataaacc tgaagcactt 240
 cttaattatt cagaaaaatg tgcaaataat gcttaattga ttttgtattt aaatgagtta 300
 aagggacagt ggataaaca acctca 326

<210> 40
 <211> 393
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (227)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (240)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (244)
 <223> a, c, g or t

<220>
 <221> unsure

15

<222> (317)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (330)
 <223> a, c, g or t

<400> 40
 cactagctca tgtagtcctc cccacaacca ggtgagacag gtgctattgt tatccacact 60
 ttacaagaag gaaacagaag tctaggaag taggtaatta acattaccca caatccgtgg 120
 gcaggaccgg gatttgaatt ggcaatgtgg ctccagtgcc tgggtgctcc acattgggag 180
 atgggtcccat caggaggtcg tctcttgaca tctccaacaa gccatcnctt tgccatgttn 240
 ctancattcc aggtagcctg agtgccccca antgaccaag gaaaagctta cccttagagg 300
 gtcttttactc ccaatgnccc ccaccttctn atcctctact ttttgttgtt taaaattcag 360
 ctgacctgtt agttgcnact ggggaaggtc tga 393

<210> 41
 <211> 477
 <212> DNA
 <213> Homo sapiens

<400> 41
 cactagctca tgtagtcctc cccacaacca ggtgagacag gtgctattgt tatccacact 60
 ttacaagaag gaaacagaag tctaggaag taggtaatta acattaccca caatccgtgg 120
 gcaggaccgg gatttgaatt ggcaatgtgg ctccagtgcc tgggtgctcc acattgggag 180
 atgggtcccat caggaggtcg tctcttgaca tctccaacaa gccatccctt tgccatgtta 240
 ctaccattcc aggtagcctg agtgccccca agtgaccaag gaaaagctta cccttagagg 300
 gtcttttactc ccaatgcccc ccaccttccc atcctctacc ttttgttgtt ttaaaattca 360
 gctgacctgt tagttgccac ctgggaaggc ctgaccactt cattctttat gcctctcata 420
 cctcagagag ctgccagggc atctctaata ctcatattt ctcaaacagt agttctc 477

<210> 42
 <211> 515
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (326)..(386)
 <223> a, c, g or t

<400> 42
 aattcatctc ttagctatag ttagtcttct actcaggagc cctttaattc aagttgtctt 60
 ttttaattatt cagtaaattc ttatagtctt tttcatattc gtcctgcatg tttctcattg 120
 aattcctgtt tttcttaata ttatgcataa cacggtattt ttttaattgca tattgtcatt 180
 atagaaacag ctgttaattg cttaacattt attttggagc tggacatctt aaatattcat 240
 ttcttagttc aaataatttc caactgatct atatagggtc tatattatct ataaataatg 300
 ctaattctca tcgccagcaa atttannnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 nnnnnnnnnn nnnnnnnnnn nnnnnnaata gccagtagcc ttgtaagtag tctagatctt 420
 aatgagaaca tctctgtata ttttaccact aagtatgaat tggctagtgg ttgtgcttta 480
 ttctactttt acactgagtg ttttaaaaca aatca 515

<210> 43
 <211> 530
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (326)..(386)
 <223> a, c, g or t

<400> 43
 aattcatctc ttagctatag ttagtctttc actcaggagc cctttaattc aagttgtctt 60
 ttttaattatt cagtaaatc ttatagtctt tttcatattc gtcctgcatg tttctcattg 120
 aattcctggt tttcttaata ttatgcataa cacggtattt ttttaattgca tattgtcatt 180
 atagaaacag ctgttaattg cttaacattt attttggagc tggacatctt aaatattcat 240
 ttcttagttc aaataatttc caactgattc atataggttc tatattatct ataaataatg 300
 ctaattctca tcgccagcaa atttannnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 nnnnnnnnnn nnnnnnnnnn nnnnnnaata gccagtagcc ttgtttgtgt ctgatcttaa 420
 tgagaacatc tctgtttatt ttaccactaa gtatgaattg gctagtgggt gtgctttatt 480
 ctactttttac actgagtgtt tttaaaacaa atcacttgag ctgctccaaa 530

<210> 44
 <211> 446
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (425)
 <223> a, c, g or t

<400> 44
 gtggtgggaa ggcaagagaa ttctgtgaaa tggactcagt gttccctgtg aaataggagg 60
 cagtgttgta agccaagagt gagcaaagag gtgggagaat cagaggtttg aggaagccag 120
 ctcaagaaga aagtgtgtag cagagctgat gagatgaaag tggcatgctt gctgggcagt 180
 gtttagagcc catctgagaa tagttataat aaatacatgg tgaaattgat ctgccctgtt 240
 gtagcacttt ctcaataaaa ctgagcagct catgccctat ctcagagcaa gaggagagtt 300
 agattcattg agttggattt ttgccagatg agtgtgataa aaagattgcc cagagtttag 360
 agttctgaaa aaagtgttat ggagtgggtg acatgagctc aaagtttgaa aaggatggga 420
 atgangaaaa gaaactagct gataga 446

<210> 45
 <211> 906
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (707)..(812)
 <223> a, c, g or t

<400> 45
 cagctcttct gtgtcaaaaa caaacaccct cctcccagcg ctgctcctgg ccggtgccc 60
 cgccctctgc caggcgttcc tcagaggaca agacctaatg agctggctgc tgccagcctg 120
 gtcctcacag ttcattcagta ggattccaga caggcatcag gctcaggagc agcgcagaga 180
 cagctgcctt ctctcttttc ccggaggcac ctgagacctg agcgcaccga gggggccggt 240
 gcatgggctg ctcccagtga gcgtgaagtt cacgcccaga agtacacccg ccaccagctg 300
 cagcagcaca ggttcgtcca gcgcaccag agaggctggg gctctctggg agtggaggag 360
 cagggtggga tgagcctgga cttgcacgca gagctctggg ctccattaag cccccgccc 420
 gtcctagctg tctcgtctgg gcacgccagt tctccctgag ctgctctcct cctggcagaa 480
 ggggggtcat aacagacca acatgcggga ttgcggtgag gtctaaacag tcaggcacag 540
 gaagctgcac agagaagatg catgggcaac agcggccatg gagaatccat gcagccccct 600
 aagaggggca gagagcctcc aagcaaaagt cattctatct caacactcac tcccctgaag 660
 actattcgtt cttgggaaat aggatacca atattgaatg tttgtgnnnn nnnnnnnnnn 720

17

```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nntaccacc acaggattac aaggagaaaa 840
agaggaaagg gatctccccg ccctctctct ttctccccct ctccaacca gggcagaaga 900
agaaaa                                           906

```

<210> 46
 <211> 289
 <212> DNA
 <213> Homo sapiens

```

<400> 46
aaacacagtc cttccatgag ttctgcaaac cttgggttgg aaaagaggct ctagtttgcc 60
ttaggtacc tggactgagc aatataaggc atgggagagg tggtttatct gtttaagggtg 120
ccatgtcttg ttataactca ctgatgagaa gaaaaaaact taaatgaaga cttcagactg 180
aatttttttt ccttgattta aaaacttaga gtgagagtta agcttagatt tagtttttct 240
aaaaccttaa aaactagaaa ccatttatta aagctagatt ttttttttc 289

```

<210> 47
 <211> 299
 <212> DNA
 <213> Homo sapiens

```

<400> 47
gggctgagct aaacacagtc cttccatgag ttctgcaaac cttgggttgg aaaagaggct 60
ctagtttgcc ttaggtacc tggactgagc aatataaggc atgggagagg tggtttatct 120
gtttaagggt ccatgtcttg ttataactca ctgatgagaa gaaaaaaact taaatgaaga 180
cttcagactg aatttttttt ccttgattta aaaacttaga gtgagagtta agcttagatt 240
tagtttttct aaaaccttaa aaactagaaa ccatttatta aagctagatt ttttttttc 299

```

<210> 48
 <211> 197
 <212> DNA
 <213> Homo sapiens

```

<400> 48
acaggcgtga gcaccatgcc tggccccaat gggatttgtt atggaacttc ataaatgtat 60
tgtaaaatcg tcatagggag aaacaaagaa ccaagaagag ccaaaatact cttgaaaaag 120
aggacaagggt gagggagttg ccctaatttg gaagctatta agatttatta taaagctata 180
ataattagac atgatac                                           197

```

<210> 49
 <211> 453
 <212> DNA
 <213> Homo sapiens

```

<400> 49
ttacaggcgt gagcaccgtg cccagcctca agtatactct tacaacacaa ttaaattcaa 60
tcttcagtaa tcccaaaatt tcattacccc tgtgaaaatg tcctggatta gcagtctcct 120
actttaagtg ttttatgaaa gaatacagtt tatttttagta taaataatat agccagactc 180
tatgaacaaa aagggtgaat aatatttacc tatagctccc atttagaagt accaaagtta 240
tgaagcacat tcattggcta ctgtcatatt tattaggatt tatgttttat cagattataa 300
gcactcttta gtgaaaaatg tttttttcct ctttgctcag aaaattgtcc aacactcctg 360
gtccagtcaa gagtgaagca aaaaactcct caatttgaat ggctttcatt tgggtccatt 420
tatttggtta cagagaagtt ttgataaaat acc                                           453

```

<210> 50

<211> 1012

<212> DNA

<213> Homo sapiens

<400> 50

```

gtaacattct atttataatt atgtccttgt tttattaatt ctctatgga tggatatatta 60
ggttatatcc atttttttgc tagtctttgt atgctccctt gaattttatt gtacatatatt 120
tcttgggtat ttgagagatt ttctggggta tacatatcta agatctgatg gatgctggga 180
tattgtgcttt gtcaactgag gttctcactc ccctggaagt gtgtgagatc agaatgccc 240
tgccctagcc cttacttata ttatgtatca gcatgattga tttgtaatag actaataagg 300
gtaaatagct gagtgtatgc cttctatact gtaattttac tttgttggtc gtctgtttgt 360
ttaattgggg acccatcttt tttcagattg ttaattttgc taaagatctt ctttgttctc 420
agagttaatt atcccttaag gaattccatg tgtttatttt tctctgttcc aaagtacga 480
ttctgtgcta aagtcataat tatgaaatca tcagtttgtt catactttaa atctatgctt 540
ctcccttggt gttgacagtc cccaaggcag gcatccatga agtcaaaagg actgaccaa 600
gtgtaattgt ccctttttac tgggttgga tttgtgctaa taccctgcaa aagcagtgg 660
ggataaactg acagcacctt gcaaagcagc aagggtggtg caccaatttg tcattattta 720
tggtaaaatt aatgggttca tttgtatttt taaatgaata aacattttaa caatttctta 780
gttttgattt ctaatagagt aactatagat cagtagatgc caactatagt gtcttctctt 840
aagagcgtga aggggcctga gactggaaag ctggagaagc accgctttta agcacatgg 900
agacgtatga atagacaaat actttattct tggtgaacat ggtcattggg aaggaaaact 960
gaggtatgtc attctattac aagatgaatc aggtgatct gcaagttgta ta 1012

```

<210> 51

<211> 268

<212> DNA

<213> Homo sapiens

<400> 51

```

gtggaaatta atgttagaat ttgtattatt tagatgaagg gaatgtagcg atgagttttg 60
taaagggaact ggtcatcgaa aggaagggga aaagatgaaa ataaaacaaa ataagaatat 120
aaaatagcca gagagattat acgatcatgt attaactcct cctgagaata aaatattata 180
ttgttatgtt tgaggctcat tttgactcag ttcttagtta agagttggct aacaaaaagt 240
atatcattgt aatgaatgct ttcactgt

```

<210> 52

<211> 581

<212> DNA

<213> Homo sapiens

<400> 52

```

gtggaaatta atgttagaat ttgtattatt tagatgaagg gaatgtagcg atgagttttg 60
taaagggaact ggtcatcgaa aggaagggga aaagatgaaa ataaaacaaa ataagaatat 120
aaaatagcca gagagattat acgatcatgt attaactcct cctgagaata aaatattata 180
ttgttatgtt tgaggctcat tttgactcag ttcttagtta agagttggct aacaaaaagt 240
atatcattgt aatgaatgct ttcactgttc ttgttcttgc tggttaaacct atattctccc 300
caggctgtgt aatccacttt tgttactctt tgctggagtc actagatgat acacaaagga 360
aattttgtgg cactaaactca gtttcgcaca tttttggcta tgaaatgtgg acagaaatta 420
ttgaaactaa tatctaaatg tagctattct ataacttcta tctagccatg ttaattttgt 480
tctctattaa gacggacaat caaagaggaa ataaacagaa catatttctc ctaatgaatt 540
caggctgggg ctaaaagttc aatattttata gatttcttct t

```

<210> 53

<211> 597

<212> DNA

<213> Homo sapiens

<400> 53

19

```

actgcatctg ctgcctttac acgggactgc aaacctggtt ttttcaacct tctgttttat 60
gggtgtgcac acccataaat ctccgtgtggc tgggttaagg gaacatacaa gcagctcttc 120
agcattaaga atgtgatggg agagattcag gtagatttga actgccatca tcaatcaaga 180
ccaaggagaa ggctgctttc caggatgtac acatggcctc tgtttgctgt tgctgttttg 240
cttcttttaa gaggtgaacc aatatatgta tgtctgtttc tactgtcact tgcagctcaa 300
cagaaccctg taatatacat gaacaagttt ctggaagtta agagagatga gaagttcacc 360
aagtcaccaa cctgactggt accatgagga attcctttac cggagaacat gctgtcacia 420
taggttaaat atatgttata cagggtccaa gaattattcat gttcaatctt agttaaaaaat 480
aaatatttat agttagttaa attaggtata gcttttattt cccacattat aattacctgt 540
attttttata cttcatgtaa catcaccaaa aatttttagta ttagataaat caaaaaa 597

```

```

<210> 54
<211> 304
<212> DNA
<213> Homo sapiens

```

```

<400> 54
gctcgagatc cctcttgtca tccaaagaga acaccaaact ggtgttagct atatttttaa 60
ataggacaaa aagtccctgc cagactgtgg agtctctcca cctggagaaa gcattcaatc 120
tctgttatgt tcatgccttt cagtaccatt cctttcgtat tttttcagtt gacatgacct 180
ttaaggttcc tccaaactaa ggttctaatt ttttttttta acttgcagtc ttactcccaa 240
caagaaattt gatataattag agctaacagt tctaagaagt ttaagaaat agtatgcaat 300
ccca 304

```

```

<210> 55
<211> 2631
<212> DNA
<213> Homo sapiens

```

```

<400> 55
caggtacaca gtgcacaaat tagatattca ttctaaaact tctaatttac agataagacc 60
gagaagaggc tagtaagtca ggtatcttaa ataattggatt cgttgaaact ggctcttcag 120
aagagggtgat tgcagaagtg caaagctggc tctgagggtta aatctttatg agaaaggaat 180
acctttactt tgagggtatta aatggctcag ctctgggata tgaaactttt taagtatctt 240
taagcaatca gtgttcaaat caaagagtga gatgcgtaat ctgacctgtt aaaatcacaa 300
aatcaggctg ggcatttagat aatgcctttc agtttaaatca ctgcgtgcct ggattctgga 360
aaatgttgct atataaaaca cataatgtat gaatagaagt atatggtaac tgacagactt 420
ttgttatata gtgtgataaa gtgaatagaa cattagaata ctaaccgcat gattttgact 480
ttggtctcag ttgttcagtt ggggccttag tttctttaca ttaaaggaga agactaaact 540
aagtttatcc tttcaaaaga ccctttacta ggtgtccttg tctacatttc caaaatattg 600
gacttggtcca tgaccaaaca ggtgggaatg aaggccatta ttttgattat ttttctcttt 660
taagaatttc cagaaatatg ttctttgtag ataaagaatt acatatttgt agagttctaa 720
gcgttcttaa aattcatatt gcccaactcc ttcttttctt aaaggagaca acagaagctg 780
cagaaatagc ctctctgtta ttattacata gcagcagctc cctgtcttta aatatttgaa 840
ctaaacacat ttacattttt aatgaattta atttacagtg tgatgtccag tattgggatt 900
gcatactatt tcttaaaact tcttagaact gttagctcta atatatcaaa tttcttggtg 960
ggagtaagac tgcaagttaa aaaaaaaaaat tagaacctta gtttgaggga accttaagg 1020
tcatgtcaac tgaaaaaata cgaaaggaat ggtactgaaa ggcatgaaca taacagagat 1080
tgaatgcttt ctccagggtg agagactcca cagtctggca gggacttttt gtcctattta 1140
aaaatatagc taacaccagt ttggtgttct ctttgatga caagagggat ctgtcgtttt 1200
aatgtcttct ctgcagccc cctcacgcga gcccctcac acctgtgagg cttctttgac 1260
gttgagcgtg caacaaccgc tgcagtcgc gggttcccaa gtgcccgccg agccagcttg 1320
caggggagtt gtgcgcggtg gctacagcct gttgatccca tttcctctct ctctagtcgc 1380
ggctagggag tggctctgcc aggacttcca aggttttttg tctcgggtac tgggtgttcgc 1440
atggctcgag tgtattgttt tcttccaggc aatctcgtgt agcgttccag cttagacact 1500
tcttgctcgt tctgtcgtct tgggctgcgt gtagtctctt gtttctgcgc tttctccacg 1560
cccttccag tttcctgtta gccgaagggg atcgtctctt ctgaacgaaa agttctcaga 1620
gcggagctga acctcccga aaatgctctt ctcttccgtg tgcgccggat gggggtgggg 1680
gtgggccaga aactgaacgc cgccgtcagg agagctgagg ggaccgcacg gccctggcgg 1740

```

20

```

aggcgggaga ggtacggtcc tcggagtggg gctgggggtg gggaaaccga cgagggggcag 1800
cccccgactg tcttgggtgc agaggggact tttattcagc tggaaccgcg cggcgaggcc 1860
caagtgtctc tggagagatt cgggggttcag gaggtggcgg gtgcacccaa ggggtgctggg 1920
aggaagctcc aggttcccat tcttccccag ggatcggcgt tgcccttctt cgcgggggta 1980
gtctagggca acggaagatg gcggcgccgg ccgggcacgg ggttcggggc tccgctcggg 2040
cagagcccac ccgctgacca actccgccgc ccccgccggg cgggtgctgtg tccccgcagg 2100
agtcggaag gatggcaggg gccggaggcc agcaccaccc tccgggcggc gctggaggag 2160
cggccgcccgg agccggcgcc gcggtcacct ccgcccgtgc ctccggcggg ccgggagagg 2220
attcgtctga cagcgaagcg gagcaagagg gacccagaa actgatccgc aaagtgtcta 2280
cctcggggca gatccggacc aagggtttca tcatgttggc caggctggtc acttctgagc 2340
tcaagtgatc cgcccacctg gcgctcccaa agtgctggga ttacagggtg gagccaccgc 2400
gcccgccga aaacaatata attgtgaagc agttctacac catgttcgta gcagcgttat 2460
tcataggagc caaaaagtgg aagcaacca actgttcaact gatggatgaa tggataaaca 2520
aaatgtggca cacacatata ataattggac attattcagc cttgaactgg agggaaattc 2580
tgacaggatc ctgtgagggtg aaaggtcgca ttttcagggtg tcagggaatc t 2631

```

<210> 56
 <211> 401
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (279)
 <223> a, c, g or t

```

<400> 56
ccttaaaaaa atttacagaa cacaaaggaa aacataaaca caaagacatg gaaaattttg 60
tcaactcctt aatggaattc tgtgatcaaa aagcaggcca gattctaatac aaaatcagggt 120
aaatttttaac cacaatcaga agtacttgta acatttcagt tgcctaact ccaatgagat 180
aacaaggcct ccaaggctac agctgaaact ctgaaaggcc ctgtgctttc tactttacat 240
ttagcgtcta atatttccta ggacagtagt tcccaaagna ggctgtacat agaactctcct 300
ggagagcctt ttaaatgcta atgccaataa ccatactctc ataaaattta ccctagaatt 360
tccctgggat ggggtgcctg gccatccagt attttttaat g 401

```

<210> 57
 <211> 859
 <212> DNA
 <213> Homo sapiens

```

<400> 57
gcacgagtta gctttgcatt atctaacca tttattttta atctgccagg aaatcctcta 60
actttccttc cttttgttt cagtaagtat caggcagctt caccatacct gagtcccttt 120
gtcttgaagc tgccacagaa aaatcttaca gcaatcattg ctgattagaa actgtttcag 180
acaatcagca tgggtgttat ttaccaaatt cccccagag tcctaggcct cttctccaga 240
aatatctgat gatgaagtga ggggagggca acggtgctac aaaacacgga acagaggtaa 300
agagaaggca ctactttctt gccatacttg taaatgattg ctttgttcaa acataaataa 360
tcttaagtcc aacaccaaact acctgttact cctacatcaa tctcattagt ggtttaagac 420
acagtactag aattttcatt ttttaaaatc ccttggccct taaaaaaatt tacagaacac 480
aaaggaaaac ataaacacaa agacatggaa aattttgtca actccttaat ggaattctgt 540
gatcaaaaag caggccagat tctaatacaa atcaggtaaa ttttaatcac aatcagaagt 600
acttgtaaca tttcagttgt cctaactcca atgagataac aaagcctcca aggctacagc 660
tgaaactctg aaaggccctg tgctttctac tttacattta gcgtctaata tttcctagga 720
cagtagttcc caaagtaggc tgtacattag aatctcctgg agagcttttt aaatgctaata 780
gccaaatacc atatctccat aaaatttacc ctagaatttc cctgggatgg ggtgcctggc 840
catccagtat tttttaatg 859

```

<210> 58

<211> 343
 <212> DNA
 <213> Homo sapiens

<400> 58
 gctcgagtgt aaacattcac tgatcttttt tcctttattg aagccacaat ttaaaaaaaaa 60
 aaaatactat aaatttcagt ttaaaattgag aagccagata tctttcaaaa tgtatccttt 120
 atgtggtaaa atagagaata acattgtttt tagttaagta aaactaaagt actgttttcta 180
 actaggtaat ctggccttcc aaacacagga gtttgaacag agagttctaa aaattagagt 240
 gtctgttctc tgtcagaacc ttctgggaag agtgtgtcaa atgagcacta ctcaggagaa 300
 atttctaagg ttttaactta gtttatactt taaactgaga ttt 343

<210> 59
 <211> 635
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (33)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (111)
 <223> a, c, g or t

<400> 59
 tcttaatgtg atttaaaata ccgggggatga agngcattca gtatctgcct ggtcaccâaa 60
 gtccaatgcg acatcccctc tctatagaga tgtattctag caaaagactt nttcatccac 120
 catctggccc cagactaaga acacatctca ctgaatgaca cataaccag tgggatgcac 180
 caaatttgct taaccatgag cacatcatct tctcataaca aaagctgaat atgaccctaa 240
 ttttatattc tgtaaaactct gttgtggaaa ttattaaaac aactgtcttc tgggtagtct 300
 gtaaacattc actgatcttt ttccctttat tgaagccaca atttaaaaaa aaaaaatact 360
 ataaatttca gtttaaatg agaagccaga tatctttcaa aatgtatcct ttatgtggta 420
 aaatagagaa taacattggt ttaggttaag taaaactaaa gtactgtttc taactaggta 480
 atctggcctt ccaaacacag gagtttgaac agagagttct aaaaattaga gtgtctgttc 540
 tctgtcagaa ccttctggga agagtgtgtc aaatgagcac tactcaggag aaatttctaa 600
 ggttttaact tagtttatac tttaaactga gattt 635

<210> 60
 <211> 474
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (335)
 <223> a, c, g or t

<400> 60
 gggaggcaag aactattttc attttatgtc ttatgaaact acagtgcata gtgacgaagt 60
 gatttgoccta aagtcacaaa gcaaaaacta ctggaaccat gtcccaagct aaagacttct 120
 cccaattata gcgttttttc ctcccatagc ctgttttcat taccttcttg tttatccatt 180
 ggctttcatg agacatgttt gctgccagtt gtgaataggt tagttcccca gaggaccat 240
 gagtaccaca caaactgcta gctgaatctt gtgagaattc taggaggtag ggctataccg 300
 gccctgaaga aatttcttga tgactgctca gtggnnttat ggaatgtagc agagtattct 360
 ctggatactt tagagttact cccttttaag agcatgatat tgacaattct ttttactagt 420
 ggaacagtga catctgaaca gcgtgcctga cctttgcaag gtttaagcaga atgc 474

<210> 61
 <211> 526
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (415)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (417)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (475)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (482)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (521)
 <223> a, c, g or t

<400> 61
 attttaaat ataattaaat attttattcc tttattatag gaagagcttt tacgagttct 60
 actgaacaac aacaaaaaat ccagtagaaa tggtggacaa aagatgtgat tatacaaac 120
 tagaaatgca agtaaacata aaaagctcaa acttacttaa aaacttaaaa tgaaatattc 180
 gtaaataaaa ctattactga gggcctataa aattttgggt taaaatgaaa tggtataact 240
 taataaatgt tagggcacia tgatgctatc tttcttacat ctttcttttt agaagtaact 300
 tatttcaatg tttctggaaa gcaatttgat aatttttata ttactacaaa aatatggtag 360
 ctaccctttg gctcaacaat ttttttagga accacaaaaa tgcagtcaaa gatgnaata 420
 aaagactgaa agcaattctt catagccttg tttatatgaa gggaaactga aaacngccta 480
 antatttaac aataggtgaa atgattagaa atgtggtata ntcaga 526

<210> 62
 <211> 164
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (143)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (148)
 <223> a, c, g or t

<400> 62
 gacatcctat acaaaaaaaaa atcgatttgt gctttattta cataaaaaata aaactatact 60

23

```

tttgataacg tcctggggcac ttccctctgc ttactcccc tcaattaaaa aatgcctaatt 120
ttaaattaaa agaaccggcg cantgcantg ttcatgccta taat 164

```

```

<210> 63
<211> 257
<212> DNA
<213> Homo sapiens

```

```

<400> 63
agcatgggtg aagctaaggt gaccttgatc aagttgcaa aacctgtttc aggtttgctt 60
aagtcaccag aacgctttga ttgagacatc ctatacaaaa aaaaaatcga tttgtgcttt 120
atttacataa aaataaaaact atacttttga taacgtcctg ggcaattccc tctgcttact 180
ccccctcaat taaaaaatgc ctaattttaa ttaaaagaac cgggccaggt gcagtgtttc 240
atgcctataa tcccagc 257

```

```

<210> 64
<211> 572
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (179)..(265)
<223> a, c, g or t

```

```

<400> 64
cacactttct cagctgctct tggttttgca aaggaagata ctgacatggt cagattaaga 60
aatcgtaaag cttctgaact actaaggaag ggaaaagagg ggcccagggc ccacatgtgt 120
gccagggtgct gatctgaggg ttttttgtga ctcatctcat ttaatgggtc cactgttcnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
ctccatatcc tgtgtccctg ggaaaggaag gggccatagt ctggagtggg ttccaggaga 360
aaagagccag agtaatctct gctcttcatt tcttaacaag aatagaagac agaataaagg 420
gcacagggat aaaggattgt taaccagact ggcaaatcag tagactaatt aaaaatcaaa 480
caccttaaaa cactgtcgct gggtaattg taaaccaaca atgaaacgtt aaatttgccc 540
agccatgagt ttgaatgatt aactgagtga gt 572

```

```

<210> 65
<211> 277
<212> DNA
<213> Homo sapiens

```

```

<400> 65
gctggctttc ggtatttata agtgcctggg aatgttctag gctctgggtc aagcctgtag 60
ggaaaaacct gcagctggct gagccacaga ggtcagggca gtctgtgatt ttcagtcagg 120
acacagaaag caagcaggag gaactggagg accctgcggc tgctgtaac aagaaataaa 180
aatggcacag atattactaa ttaagcacta atcccagagg cggcgagctt gtggccttcc 240
tgttctcctc ttaaaagcaa gcaagggccg ggtgtgg 277

```

```

<210> 66
<211> 452
<212> DNA
<213> Homo sapiens

```

```

<400> 66
cccaggggat gatccaaag cattttccca ggggtccttc gttgcagggt gggtctcagt 60
gtccttgcaa tgggcatcag agaaaagcgc tgttctacag ccagggtgtgt cctcggcaag 120

```

24

```

ggggtcaggg tatggagttt atgtgagggg ttaaggattt tggctcaggg cctgggctgg 180
ctttcgggtat ttatcagtgc ctgggaatgt tctaggctct ggttcaagcc tgtagggaaa 240
aacctgcagc tggctgagcc acagaggtca gggcagtcct tgattttcag tcaggacaca 300
gaaagcaagc aggaggaact ggaggacct gcggctgcct gtaacaagaa ataaaaatgg 360
cacagatatt actaattaag cactaatccc agaggcggcg agcttgtggc cttcctgttc 420
tcctcttaaa agcaagcaag ggccgggtgt gg                                     452

```

<210> 67

<211> 283

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (274)

<223> a, c, g or t

<400> 67

```

ggaataattc agcactttaa tgtgttattt aattctcaca gaagcccat tttacataaa 60
aatgaaattg aatggattat gagaatattg attattgatt ggtaagtagt aacattattt 120
tttcaagaac agcaacctaa aatactcata cagttagctc taacaatgtt tacaagtctt 180
aaaactattc ctgcaaattg ttgtattaca taaatgttat tgactcctca accatggttt 240
tttaaagtaa tatttgtaa ttataaagta aganaataca agc                                     283

```

<210> 68

<211> 432

<212> DNA

<213> Homo sapiens

<400> 68

```

ggaataattc agcactttaa tgtgttattt aattctcaca gaagcccat tttacataaa 60
aatgaaattg aatggattat gagaatattg attattgatt ggtaagtagt aacattattt 120
tttcaagaac agcaacctaa aatactcata cagttagctc taacaatgtt tacaagtctt 180
aaaactattc ctgcaaattg ttgtattaca taaatgttat tgactcctca accatggttt 240
tttaaagtaa tatttgtaa ttataaagta agaaaataca agccgggcat gatggcacat 300
gcctgtagtc ccatctactg gggaggctga gtcaggagga ttgtttgagc ctggagtttg 360
aggctacagt gagctatgat cacattattg cacgttagcc tgggtaacac aatgagacct 420
tgtctcttta ac                                     432

```

<210> 69

<211> 516

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (425)

<223> a, c, g or t

<220>

<221> unsure

<222> (475)

<223> a, c, g or t

<400> 69

```

ctttttctta attaaaaatc ttaaagcctt ttcccttggc tgtcctctga agacagtgtg 60
aatcttcttc aggcctgctt ttcctaattt tatacattat tgctctaact tatttttcta 120
cttattattt tattttctat ttaataaaat acaaactaca ttgcttgaat tgtgttgat 180

```


25

```

ctgcaaaaca atatggatac aaatacggat tttttagcta ttttcatttg ttcttttcta 240
cattatactt cttgaagctt ctgttttatt cagtttgtgt agaggtgaat gccctactga 300
agaatctgtt tttcaaagat tatccaagaâ aatatttttt gagagaattc tagtggattt 360
aattgatgaa gacatggtaa gagaaactgt tggaagatac ttgaaagaaa gtcattaagt 420
gaganaaaaa tggagaacta aaatgtggag actcacgaag agcagagtga gcttnaagaa 480
taaagactgg aaacctgtgt ccttaatgca tttact 516

```

<210> 70

<211> 52

<212> DNA

<213> Homo sapiens

<400> 70

```

cattgggtta atatacctga gcacagttta tgaacctttg tcctcttcta tt 52

```

<210> 71

<211> 422

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (311)

<223> a, c, g or t

<220>

<221> unsure

<222> (386)

<223> a, c, g or t

<400> 71

```

ggggaagata cttgagcaca tttatagacc catgataagg agctataaaa ataatgaggt 60
taagatgctg acaactatth atgcaaatac cagagaaatag ttagctttga acagaagggc 120
acccatctct tctctaatat tggaaacagg tggaaaaacc acctgggctc tcagacagat 180
gtctttgttt ttaaataatth cagaaaatga ggtagggagg gactgaccaa gggcagcgag 240
ttttatgaat gctgttcctg gtctcagcag cgctttcctc ttccctcact gacaactgca 300
gggcccaagt ngggaggaag aacagtgtgt gcctgctggg ctcagcatct gctccagtga 360
gcaacacggg ggtgactggg ggtctnctga atgttaaata taaaggaagt tccttttccc 420
tc 422

```

<210> 72

<211> 521

<212> DNA

<213> Homo sapiens

<400> 72

```

ggggaagata cttgagcaca tttatagacc catgataagg agctataaaa ataatgaggt 60
taagatgctg acaactatth atgcaaatac cagagaaatag ttagctttga acagaagggc 120
acccatctct tctctaatat tggaaacagg tggaaaaacc acctgggctc tcagacagat 180
gtctttgttt ttaaataatth cagaaaatga ggtagggagg gactgaccaa gggcagcgag 240
ttttatgaat gctgttcctg gtctcagcag cgctttcctc ttccctcact gacaactgca 300
gggcccaagt ggggaggaag aacagtgtgt gcctgctggg ctcagcatct gctccagtga 360
gcaacacggg ggtgactggg ggtctgctga atgttaaata taaaggaagt tccttttccc 420
tcttagagaa gctcatagcc aaactgaaaa gcggaggaga gataaaatga ataacctgat 480
tggaagaact gtctgcaatg atccctcagt gcaaccccat g 521

```

<210> 73

26

<211> 140
 <212> DNA
 <213> Homo sapiens

<400> 73
 ggatatttgg ttactttgca gcctagaaat tatttcagag aatcctaatt gctgacattg 60
 catatttggt cagtttgag tctggttggt agattatcaa agaaaagtcc tgctgatatg 120
 taagcatcaa atagaaactt 140

<210> 74
 <211> 101
 <212> DNA
 <213> Homo sapiens

<400> 74
 aagctattaa aggctgtccg ttaaggatct ggcttcaaac tgcctttcca ccttcattct 60
 actatttcct ctattaaaat atgctttgtg tttaagcaa a 101

<210> 75
 <211> 422
 <212> DNA
 <213> Homo sapiens

<400> 75
 aagctattaa aggctgtccg ttaaggatct ggcttcaaac tgcctttcca ccttcattct 60
 actatttcct ctattaaaat atgctttgtg tttaagcaa attgttaatt tttttttttt 120
 tttaagatgg agtctcgctc ttgttaccga agctggagtg cagtggcccg atctcagctc 180
 actgcaacct ctgcctcctg ggttcaagca cttctcctgc ctcagcctcc cgagtagcta 240
 ggactaagtc atgtgccact atgccagct aattttttaa atttttttgt agagatgggg 300
 tctcactgtg ttaccagcgc tggctctgca gtcttggcct gaagtgattc tctcaccttg 360
 gccccccaaa gtgctggcat tataggcatg agccatggtg cctgtcccta ttcttaattg 420
 ca 422

<210> 76
 <211> 253
 <212> DNA
 <213> Homo sapiens

<400> 76
 cacacctcat ctccttgaca ggaagacatc ttttttctg tggagcctgt ggaatttatc 60
 actttctatt tctcttgggt gggaaaatct tctcggcatc tagctaggca tggacagata 120
 ctgttgggtg atgatgccac tgaagagccg tccttagtgt cacgtggtgc tggctctgagg 180
 tcacggtcca ttggtgtcca ttggcttctc aaggccaata cccagtcccg gggctaattt 240
 ctactactga gag 253

<210> 77
 <211> 493
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (199)
 <223> a, c, g or t

<220>
 <221> unsure

<222> (202)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (208)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (211)
 <223> a, c, g or t

<400> 77
 tctgtctgtt cagggaaacat tctgcggcag ttaaacagca gccttcccca ttaagtccctg 60
 gcaacacagg aaaggtagat gcttttcagt aacctttccc tgtaggactc tttcagagcc 120
 aagaacataa ggtgtgaccc atctggacta aaaaaaataa agcagaattg tatcaattgc 180
 tactcctttt tattcccanc tngtttttct natttttttt ttttaattccc atcttgtaag 240
 agaattccca gggagccttt ttgagagaaa gttcattgga tttatttttt taatttttat 300
 gccatttctt gtaaaagcaa actgctctag ttggatgcca ggtatacata aatgtattga 360
 taatatccag tctcttgggg aactctagga gtatttgctt aagacacatc ttggggttcc 420
 cttacactct ttctaagatt tacaggagaa ggagagtctt actgtctttt ctagtcttat 480
 gaaagtgata acc 493

<210> 78
 <211> 652
 <212> DNA
 <213> Homo sapiens

<400> 78
 tctgtctgtt cagggaaacat tctgcggcag ttaaacagca gccttcccca ttaagtccctg 60
 gcaacacagg aaaggtagat gcttttcagt aacctttccc tgtaggactc tttcagagcc 120
 aagaacataa ggtgtgaccc atctggacta aaaaaaataa agcagaattg tatcaattgc 180
 tactcctttt tattcccato ttgttttctt attttttttt aattcccatc ttgtaagaga 240
 attcccaggg agcctttttg agagaaagtt cattggattt atttttttaa tttttatgcc 300
 atttcttgta aaagcaaact gctctagttg gatgccagg atacataaat gtattgataa 360
 tatccagtct cttggggaac tctaggagta ttgtcttaag acacatcttt gggttccctt 420
 acactctttc taagattttac aggagaagga gagtcttact gtcttttcta gtcttatgaa 480
 agtgataacc gactgggcgc agtggctcac gcctgtgatc ccagtacttt gggagggtcta 540
 ggtggttagc tagcttgagg ctaggagttt aagaccagcc tgggaaacat agactccctt 600
 tccattttta aaaaaaaaaa aaaaactcga gactagttct ctctctctct cc 652

<210> 79
 <211> 591
 <212> DNA
 <213> Homo sapiens

<400> 79
 tgcattgtgga agagatatcc caggaatctg atcttgagaa cttgaacata atgttaattgt 60
 acgtgctata ggcttatagg ctccatgaag caaccttctg ttagatcaag gcaaaaaaaaa 120
 aggtctacca ttctctactc catttccatg ccgtaaaaag ttttgtttgc cactttgaaa 180
 tctgcaatga atctagagca gtagcatcaa tactttccta acactggatg gatactattc 240
 acagcatccc ccttcctcat cgtcaccggc atcacttttc tcattaccac catccccatc 300
 actagcatct gtagcacact tagtctacaa agagctttca ttcacctgac cttcttagaa 360
 caagataatt atcaactttt ggtgctggac cgagtgtttg gacacttcat cttgcagtga 420
 ttttgtgggg gtaaatagag cagcattatt tgcacaactc ccaacaacac agtgtttgct 480
 acataaggag tgcttgataa atgtggaatt gattaatgta aataaggaaa ctaaagctta 540
 ggagaagtgc tgttggtttc tcagtatcag gaagaaagga attgcagaca c 591

<210> 80
 <211> 160
 <212> DNA
 <213> Homo sapiens

<400> 80
 gggggcagaat atctgaagag atcatggcctt gaaaacttac taaatttgat gaaaaatggt 60
 gatcttcaca ttcaagacgt tcagtgaact ccatatagga gaaattcaag agatccacaa 120
 ttagacatat gctactcaaa ctgtcaagag acagagacaa 160

<210> 81
 <211> 731
 <212> DNA
 <213> Homo sapiens

<400> 81
 gcagacagcc cggcgaaccg cgcaatgcgc tttcttctgc ctgcagcaga gaaaaggaaa 60
 gaaaactccg caggggctcc gttggcttct ccacgagtga caaccatggt ttcccatgat 120
 agacagaccg gagccctgct cctttgcgat ccgccgaggg ctgcagagag catcctcatc 180
 catttgggca cccctgcccc ggaagagccc gggccatccc cttccggga cgtggatcct 240
 ctaagagggtg aattttcttc ggtggattcc gatttgctcc gtctgaccag cctaggcaat 300
 ccagcaatcg cgggtgggtaa ccaagttgcc gcttgggcac acatggcttc acgccggctc 360
 cgcctcacca gcaagcgcca ttcccagagg agaaaatgag aactgagtg ggactcaggg 420
 attgctccag gccacacagt cagcaggagg caaagcccag attcaaatgc agattactca 480
 gctccacaat ccacatcctc acaggagggt gcactccttg cccaagcgtc agacaggagc 540
 aaagagaaag aaggcaacca gctggctact ttcttccctt cttggatgcc tccaacaggg 600
 tgagaaggac taaacaaatg accaagtgtc atcccatttt ggacatactt aaaacacccc 660
 atggaatttt tattctgact ttcttctgcc tgtgtggcat ttatgtttaa ataaaagaga 720
 attcaactcg t 731

<210> 82
 <211> 666
 <212> DNA
 <213> Homo sapiens

<400> 82
 cagtgtagca ctgtaattta tttcatttct tgactaatta ttcaagccct tgataaacia 60
 tgggttatggg atgacttacg tgtagctctc aagttctaaa taatgttaag ttttagcagat 120
 aaggcagttt atcacagtgt ccgttcactc agacagcata agtatgtgtt gataaaataa 180
 tcttaataac aagaacttta gtaaaagaaat aagccacttc attaacattg taaaatagtt 240
 ttaagatata aagtatgaaa ggaattttac agtgtatata ttttctgact ttccaattag 300
 caattataaa tttttattga caatcttatt ttgaaaaccc cggagttttc aaatattctg 360
 catttatgtt gaccatttta ccaagatgat aaaacatgca ttattttctg ccattttata 420
 atttttacag gggggaacag cgaagccaga tgatttatta gttattgccg gtgaaaatac 480
 agagatcctt tgaaacattt gtctctccta gaattctcat caaacatat gcttctaaca 540
 cagcacttaa cagtcatggg gagtatgtgg gaataacaga gactcgcttc cctggccaaa 600
 accacacata gaccacaca cttgaaaaat aaggaaataa gatcatctga gtatggagat 660
 tcctca 666

<210> 83
 <211> 673
 <212> DNA
 <213> Homo sapiens

<400> 83
 cagtgtagca ctgtaattta tttcatttct tgactaatta ttcaagccct tgataaacia 60
 tgggttatggg atgacttacg tgtagctctc aagttctaaa taatgttaag ttttagcagat 120

29

```

aaggcagttt atcacagtgt ccgttcactc agacagcata agtatgtggt gataaaataa 180
tcttaaatac aagaacttta gtaaagaaat aagccacttc attaacattg taaaatagtt 240
ttaagatata aagtatgaaa ggaatttttac agtgtataca ttttctgact ttccaattag 300
caattataaa tttttattga caatcttatt ttgaaaaccc cggagttttc aaatattctg 360
catttatgtt gaccatttta ccaagatgat aaaacatgca ttattttctc cattttataa 420
tttttacagg gggaacagcg aagccagatg atttattagt tattgccggt gaaaatacag 480
agatcctttg aaacatttgt ctctcctaga attctcatca aaccatattg ttctaacaca 540
gcacttaaca gtcatgggga gtatgtggga ataacagaga ctgcgttccc tgccaaaacc 600
acacatagac ccacacactt gaaaaataag gaaataagat catctgagta tggagattcc 660
tcaaaaatta aaa

```

673

```

<210> 84
<211> 488
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (392)..(435)
<223> a, c, g or t

```

```

<400> 84
cctgtgaaaa tgtataatgt gtaggttatt ctaaaggcat gagccaccgt gcccggccaa 60
gaaaaggaca tctttttcta atttaaacag aagcagcgaa gtcctagtgg tagccctgat 120
tagcaatatg gaaaatttcc aagtacatta ttgcttggtg cataccttac agaaggaaag 180
aagaatgaga gaggcataata ttagagagtt gtaactgcct attgtttaag gatagaataa 240
taaatactca tcttttagtat ttactaaaga tgaagttgct caggacttaa gtggcggcag 300
tctggtgtaa tggtaaggcg gcacatcggc tctgcagtca gatggcctct cttcttctct 360
aactggtcac cttatgcaag ctgttgcaac cnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnntgtag ggtggcaagg ttatacatat tataagggtta tgcattattga 480
tgtaattct

```

488

```

<210> 85
<211> 368
<212> DNA
<213> Homo sapiens

```

```

<400> 85
ctttatatgg ttctgattta tgagaaaaca cataccaaat tttgatgacc attattaact 60
attattgtct atgctgcttt ttcattcctt agaaacaacc taaaaatctt ggactgtatt 120
tttttaaatg ctaaagtagg attcagaaaa cagatttttg tcatattgtc tttgaaacct 180
cattataaat catttagctt ttgctctact tactttcagg tttgccataa agagcacaag 240
agataatata tatgaaagtg atttatactt ttgttaagag ttttggtcag tgtctaata 300
tattacagcc ttttgctga ctcagcttgg caatctagtc tgtaacttc actctaagta 360
ataatatt

```

368

```

<210> 86
<211> 133
<212> DNA
<213> Homo sapiens

```

```

<400> 86
gttacagcat tatttaacag tgaaatgttg ttctttatat taaattgtgt cttcctgtct 60
ctatagtga tacaataga ccttgtagac acagaatttt tgctattcga aacttttatt 120
gaaaagtttt ctt

```

133

```

<210> 87

```

<211> 626
<212> DNA
<213> Homo sapiens

<400> 87
gaccgctcta attaaatatt ttaagggttac agcattatatt aacagtgaaa tgttggttctt 60
tatattaaat tgtgtcttcc tgtctctata gtgcatatac atagaccttg tgaccacaga 120
atTTTTgcta ttcgaaactt ttattgaaaa gttttcttag cctaggcaac acagcgagac 180
ctagtctcta caaaaagatt tagccgggca tgggtgcatc tgcctgtagc ttcagcttct 240
tgggaggctg aggcaggagg gtcacttgag cccgggagtt tgaggcacag tgagctgtaa 300
tcataccatt gcatgggtgca ctccctctgg gtacctgatg agaccgtgtc tctaaaataa 360
gaaaataaaa taaagggtgt gggatttgtt ttttcagtag gcaggcggtt cacggaatat 420
gggacatcag tgtgcaatct aagtttctag gttttctttt ttaggttttc ttaaaaaaag 480
atgttccctc aagtaactct taatagaact aatagtactc tcaattgttt ttttcttaca 540
gggtctatat ttacgtgcct aacagtagct ctgggatttt atcgctgtg gatctaataa 600
agtgtctatt taaagtgtaa taaaaa 626

<210> 88
<211> 380
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (372)
<223> a, c, g or t

<400> 88
tgtggccaca tcagtaagtc ctgtccgata ggatatatgc aaaagtgtca actatccact 60
tccatgaatc tccttaaaag atagtgtagt cctttgccct tcctcttcat cctctctcta 120
gttgctgcct aaatatgggc atggtggccg gagctcccac tgcctggaac cctgaggaca 180
agggctgcat cctactaggg aggcagagct atgagctaga cgcaatgtgg cccctggggg 240
ctctttgcag aacagccact atcccagccc ttctagatgg ggaaagcgag gccctgagaa 300
gtgatgagaa tcagtggcaa agtcagatgt accacttcag tcacacactc acattttttt 360
gctttgttcc tntttttttt 380

<210> 89
<211> 493
<212> DNA
<213> Homo sapiens

<400> 89
ttctggacct ccatgttaaa ttcttggttt gaggcaggga aagatgaaaa cttacttgca 60
gtgtagttag ttagagaga gaaaacagtg gctgtagtta ggaacaagtg aatgttaaca 120
agtttgcttc tcaggggcat tgggttaaaca acttcttaac tggccagggt ccagcacgtt 180
aatcattaac ctagggtga gcatctgctg cctgatgtat ccagaattag tttatcatta 240
cctctaacga ccatctttta tggttccgaa gagcctctat gcagtctctt atcaccgcca 300
tgcctaactc tcatttaccg ggagcagtg gctgatgttt cttagttaga ccagagtaag 360
aagtttatgg tcagttgatg aatttttaat tataactgtt taaaaagaag acgatgacta 420
tgaacagcag ctcaactcgt gcaatctttg gacagtactt cgaagtgacc cactttccca 480
tttaactctt ggg 493

<210> 90
<211> 1119
<212> DNA
<213> Homo sapiens

<400> 90

```

ttctggacct ccatgttaaa ttcttggttt gaggcaggga aagatgaaaa cttacttgca 60
gtgtagttag tgtagagaga gaaaacagtg gctgtagtta ggaacaagtg aatgttaaca 120
agtttgcttc tcaggggcat tgggttaaca acttcttaac tggccagggt ccagcacgtt 180
aatcattaac ctagggctga gcatctgctg cctgatgtat ccagaattag tttatcatta 240
cctctaacga ccatctttta tggttccgaa gagcctctat gcagtctctt atcaccgcca 300
tgcctaactc tcattttaccg ggagcagtggt gctgatgttt cttagttaga ccagagtaag 360
aagtttatgg tcagttgatg aatttttaat tataactgtt taaaaagaag acgatgacta 420
tgaacagcag ctcaactcgt gcaatctttg gacagtactt cgaagtgacc cactttccca 480
tttaactctt gggaagcctg ggttgccctg ttttcgactt tggagggtccg tgggctagat 540
tcagagtgcc ctggcaggct ggcttgggtt tgaggctgtg gctgcagcct ccgcaacacc 600
ctatctcagc acctgggaac tggcccttgg taccogattc tttcttcttt gtgtgtgtgt 660
gtaaatcatt ttcatttttt ctaatgatca aagtatacat taaaataaat gaaagcaata 720
caagtcctatg tgtatggtag aaaatctgga caatactaaa aatgtacaga aatggctttt 780
aaagattaat tttcaaccta taaactaagc tacttttcat tttagtgtct ttttaaaaaac 840
agctttttaa aacattttta agggctcatc atgttcaaga atggagggaat gtttggctac 900
aaggccttca gtatgactct atcctatagc tggagggtta ataatacaat atattaaagc 960
ttttctaagc ctccagaagg gtttgtctgg gtcttattta ctataacagg caagttaaag 1020
aaacttgagt ttaatttata tttcagttca ctttttttag acaacaagtg caatttgggc 1080
tttatttatg gaaggagaga gttgtccttc tccccgga 1119

```

<210> 91

<211> 455

<212> DNA

<213> Homo sapiens

<400> 91

```

gcactccagc ctgggcgtcg ataaatggca ataaggagg cgtgcctgcc gcaagggttt 60
tgtgaaagct ataagaacac actccctaca aatttatacc gacacaccac agacttagag 120
gaaaagggtt cccaggccct tcccaaggcc ctgaagtga ctttctaagc caaacagacg 180
ggacatgtgg atggaagtc cacttctcaa agaaagtctg aagcaagctc aggaaacttc 240
tggagctttc tggagctgca cggaaaagctg tggatatgtg gccccatgac gtgggtctct 300
gaacttgcag agacttgacg tatggcaca aaattgcaga tggaaaagag gaaaccacag 360
ctttcacgct aatgaacagt gtttcttaca aagagttacc tggcttctag atctgtgatc 420
atgaattcca gtaaaggcaa aaaaaaaaaa aaggg 455

```

<210> 92

<211> 891

<212> DNA

<213> Homo sapiens

<400> 92

```

gaagtctatt atagcaatta gtttgcttta aatatgtaat ttatattaat ggccttatac 60
catccttatt ttgcaactaa cttttcattt aatattatat gcaagacatt tctattcctg 120
taagtatagc tctgcctgca cagaattggt ggattaatga tgtagatttt aaatgtcagt 180
ggatgtagca gaattgcttg catcaattca cctccataa aaggaccat ttctccaaac 240
ccttgccaac agtaagtggc atcaatatcg ataggttgtt ttgttttgtt ttaaccagtc 300
tcgtgattga aagtaccttg ttttcacttg aatttccctg attacgtaca aaatcaaaaca 360
tttccatggt tattggccat ttgtagatct ctaccgtaca ttgcctatta gtgttatggt 420
ggcctagttt tctgtgggt tgtttatctt ttaggattgc tcccctaaac aaaaacaaaa 480
aggcagttat tccaatgatc aacaaatatt cttttctcca ccagtttaaa atgtcagctt 540
tcaactcatc acatgctaag ttgtctatct ctgtcctgtg gatccagggtg ttggttctac 600
caaccacact cctcacgctg tcatttgggt ctttgcattg tgttttgcca ttcattggga 660
agttcctttt ccttctccoc cctcttcagt gtcttatact tgggtgtttt ctgttccagc 720
tgaatcaaga attagtttat taaattccag taaaattgtg gtgttctttt gactaggatt 780
gcatttggtt acacattaat ttgagaataa tagacctctt tataattaat ataatgtttc 840
tgattagtgt tatgataggc ctcatcttat taaagtcttt tatattcttg g 891

```

<210> 93

32

<211> 278
 <212> DNA
 <213> Homo sapiens

<400> 93
 aatatccaaa tccatgttga accagtatct ttgaagattg accttcagga taaactaata 60
 gcacaatctt catgcccttt cattcataac taactgaaaa ggttgacttt tgtaccagga 120
 acccaacgaa gaaaacttgt atcttgctcag gggttggtaac ctggctgcca ttgactgaga 180
 ccaaaatc ccaacagttt gtcctcagct gctaagctgc tgtggttaga atcaaacgta 240
 gaggttctgg cctgggtcgg tggctcatgc ctgtaatc 278

<210> 94
 <211> 274
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (95)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (194)
 <223> a, c, g or t

<400> 94
 gattaccaga ttttattttt aaaatttttag caatatcggt cttaatatta gccaatcac 60
 tgcctatgga tgcagacca ttttccctg cacancctt gtagagacct gcctgggtgct 120
 cagaagaaga aagatngaatt ttgctgttcc caggaaatgc tgcacattgt ccatttacca 180
 gcatcttata gaanatataa atatgaatct acaaattctc ttggatttaa taatgtaact 240
 tatatttatc ataaggtggc tattccagat catg 274

<210> 95
 <211> 130
 <212> DNA
 <213> Homo sapiens

<400> 95
 cagataccac tctaggtgat gactgccagt ctgtgcttac agcccaaacc tctcctgagc 60
 accaaccat atgccacgg tgcagagaca gcacaacca gtgtcaagga gcctggcttt 120
 taagagtc 130

<210> 96
 <211> 1100
 <212> DNA
 <213> Homo sapiens

<400> 96
 gtggccactg cccttggat gaataataat cactgaca tacaactaag aagttatgga 60
 atacattagg aatgctgagg gcacatggaa aacagtgacc cattctacct agtgggggtt 120
 taaaatactt atttttaatg tttaatgctt tagggaagaa agcagggaga tgaaacatga 180
 aagatgaaca ggaaatggta ggagatctt atgaaggtag aagagacagg gctttgggaa 240
 tggatacccc caggttaact cccagatttc tggcttaggc aactgagtgg caccactgtc 300
 agagcctaga aatacaggct tgaaaggaga gatgctaagt gtagctttgt tggctctttg 360
 ataaatatgc gacctgcaag tggagctatc caggaaatac aagtcaaaag acccaagtcc 420
 tcttgagagt ttcctctgag ccataatatg tttcctttct ttttctttt tttttttctt 480
 ttgagacaaa gtctctctgt cgcccaggct ggagtgatgc aatggcacaa tcacagctca 540

cgaggagcttc	gacttcctcg	gctcaggtga	tcttcctacc	tcagcctcct	gaatagctgg	600
gacaggtgcg	caccaccaca	tctagctact	ttttgtattt	ttttagtagg	tgggggtctcg	660
ctatgttgcc	caggcagctc	ttcaactcct	gaggctcagg	tgatctgccc	gcctcggcct	720
cccaaagtac	tgggatcaca	ggcacaagcc	actgctcctg	gccatatatg	gtgttattta	780
atcctcacaa	caaccctatt	attatgcctc	cctttaacag	ctgaggaaac	tgaggcacag	840
agaaattaca	taacttgccc	aagattacat	gactcttaaa	agccaggctc	cttgacactg	900
ggttgtgctg	tctctgcacc	gtgggcatat	gggttggtgc	tcaggagagg	tttgggctgt	960
aagcacagac	tggcagtcac	cacctagagt	ggatctgaa	gcctcaagag	gagacaagat	1020
cacatggaac	gccacggaca	gaaccatgtg	gagcaccatt	ctcatctagg	taggagtcct	1080
caaagaaggt	taaaaagaaa					1100

<210> 97

<211> 591

<212> DNA

<213> Homo sapiens

<400> 97

cgatgttttg	gatatgtttg	ctagttataa	attaaataac	tatagttatc	cagtttttagt	60
tttgtatgct	actctcttcc	ctcatcatal	gatattttta	aaatctagtt	cagtgtttct	120
gatatatgat	ccaaatagta	ttaatattat	taatgtgttg	aaataaacac	actaatcac	180
cttagcacac	agtatacaca	ctaaaagtat	taatatgttg	agtgtgtata	tttctataaa	240
cactaataat	atagaaatat	acacactaat	aatattaata	ttattttatt	atttttgcct	300
cttcattttt	tgttgatcat	caactcatcc	ttagttaacct	ccaccatcat	cacaaatctt	360
ttaatattac	taaaccctta	ccttccttgg	ttataaatta	aaattaaaca	caacttttgt	420
ctctagagat	gcagatatag	tctgtgaagc	tgctttgatg	gcagtgattg	tgaatttcct	480
ctgattgggt	caggtttggg	taaatttctt	tcagtttttt	tactctagtt	cctactacca	540
atztatagtt	agcttaggac	ttggacacca	gaatctaagt	ctatgagaaa	t	591

<210> 98

<211> 1550

<212> DNA

<213> Homo sapiens

<400> 98

gatcttacat	ggcttatattg	taacctgcag	tattgaccat	tgcccccttat	aatttatagg	60
taaattctgt	tgatcagcat	ttttaacagc	tcaatcgatg	tttttgatat	gtttgctagt	120
tataaattaa	ataactatta	gttatccagt	tttagttttg	tatgctactc	tcttccctca	180
tcatatgata	ttttaaaaaat	ctagtttcagt	gtttctgata	tatgatccaa	atagtattaa	240
tattattaat	gtgttgaaat	aaacacacta	atacacctta	gcacacagta	tacacactaa	300
aagtattaat	attgttcagt	gtgtatattt	ctataaacac	tcaataatat	agaaatatac	360
acactaataa	tattaatatt	attttattat	ttttgcctct	tcattttttg	ttgatcatca	420
actcatcctt	agttacctcc	accatcatca	caaatctttt	aatattacta	aacccttacc	480
ttccttggtt	ataaattaaa	attaaacaca	actttgtctc	tagagatgca	gatatagtct	540
tgtgaagctg	ccttgatggc	catgtgattg	tgaacattcc	tctgattggt	tcaggtttgg	600
gtaaatttct	ttcagttttt	ttactctagt	tcctactacc	aatttatagt	tagcttagga	660
cttggacacc	agaatctaag	ttctatagag	aaatggactg	agtctgtcct	gttcacagct	720
agatcttgaa	cacccaagaa	tataatacct	gatgcaaagt	agttggtact	cagttagatat	780
ttgttgaaatg	aaaaatgtcc	aaatcaaaga	aaccacagtc	tgatgcccct	atattcctat	840
atacaaaatt	gtacattata	cttaatatatac	agaagtgtat	attaaacctta	aatgtttctaa	900
tactattttt	acatctacaa	cataaaaaag	aataatgtag	gctcaaataat	cagataaatc	960
taggttgaga	ttgtggcctc	atcattttact	tgaagtgtgt	tcttgggcat	attagtggc	1020
attctaagtt	tcagttccct	cttctcttga	ttatataata	attactacat	ggaattgcta	1080
tgggagatta	atacaataaa	agctcatagt	actgtggctg	tctaactttt	tagctgtcat	1140
tattctaaca	gttattacta	tcttattctc	aactgttttt	aaatagtatc	ttgtgttttt	1200
ttaactttat	gtccatttta	ctgttcactc	ttatgagcca	cagagtctgg	aatccagcct	1260
tggttctctc	agaactattg	attttctatg	ttcttgttgg	aacaattttc	tgctttagaa	1320
aatctgcac	agtcttcttc	tttacagatt	ttccctcttt	attgtaaaga	tctttaatcc	1380
atggaattta	ttacaaatta	atgattaatg	gactggctct	gagtgaataa	ttcagcagct	1440
gaactaaggc	tgtcttaaat	agcaccaata	taagtgaatt	caggtaaacac	acaattgggt	1500

cttttgctct gctgcagatg atagctcttt acctttcgtg gtttttccat 1550

<210> 99
 <211> 535
 <212> DNA
 <213> Homo sapiens

<400> 99
 tattttccaat atcatctcac tttatatattt attcatgtaa actaagtttc taaagtggaa 60
 attttagaat ttcccttctg cttctgaaaa cacttcagtg gcttttcatt ggcccaaact 120
 ttttgggggt agtattcaaa agtttcatga tttggccctc atttgccttt ctgatgtcat 180
 catatgtcac tctctcccat atactttaat gcaagttttg tgatctctga gtacatgtca 240
 aactttact ttaactccac ttgtcatttg tgttatgaag actggaagcc ttccttttcc 300
 ccaggccttg gtgaagccaa gtgctttaca tacctggaat gtctttgtca cagtaatttt 360
 cagttagttt gtaattgtct atttaattaa cttagtcccc tctactagatg tgagtacttt 420
 gaaggcagta atttttttta aacaatgggt gcttaacatt caataatagc ttgtagattg 480
 tgggtgattt atatatattg gcagtttttt aatgttttat tttgaaactg gagat 535

<210> 100
 <211> 493
 <212> DNA
 <213> Homo sapiens

<400> 100
 acatcccttt gctacacttt tcatgctcat gttttacttg gtgtgtgatt tcctttctgc 60
 ttcttttacta aaccgtggac ttcttgaggg cagaggctag atcttggttca tctttggagc 120
 catagtctcc attagttgga tgaatgagta cgtgaatgaa tgcttgaaatg aatggagtgg 180
 aatgaaccct gtctccccag ttctatgtcc acctcttatt cattctgtga ctttgggtag 240
 gacatttaac catagttaat aagaatcatc acaactgttt gcctttctca catactggtt 300
 tcaaaatcaa taaagtattt aaggcaatgc tttaaaaacc atacagcact gtacaaatat 360
 gctatcccat ggtaagtag aatttagtgg gaaaattgat accatcaaca tttgattgca 420
 atgattttat ctaatagaaa attaatcttt cctgggcaca gtgggcttca cacctgtaaa 480
 tcccagcact tcg 493

<210> 101
 <211> 843
 <212> DNA
 <213> Homo sapiens

<400> 101
 ggccgaaata caatttaatt aaaatactta ttttcttatt agaaagctgc ctctcaatgg 60
 cacctactgc tacattttaca tagtaaccca aaattgcagt tgcttagcag ggagagaatc 120
 acagtgtctg atattattta tactttttct tccaaaacga tttgaggaag tactgtgctg 180
 gccattgttt acatcatatt aggagatctg gatgtcactt tcttttccca tatcctcgat 240
 ttcttcactt tttaaaatgt catgtgtttt tgtaagtttt cttaaatcct ttggaaatgt 300
 gatgacggtg aaaaatccct gaaggcttct atacacctgt tggcatggaa tattttgcaa 360
 cccgtttctt ccctacaaac agaagagaca actaaatacg gtttgatcta catctgcaag 420
 agcctagcca ttcagtatta aaaagtgatg gccttggttg acggtaccac acctgaagac 480
 ctatgccctt tccttcacac tccctacttc tgcatttctt ccctcctgaa cgtctatcaa 540
 gtggaccata tgaaattgcc agtattcaac tgttttttat cttaaaaggt gacaattcta 600
 tatcattcaa cctaaattaa tgtctcaaga acataacctt tgtttctatt attgtgacct 660
 tacttttaac catcctagag ctctttaacc tgttcacact ggatttcaag gatcttaagt 720
 tgtttacta cataatcact atcacacttc agaaacattt tagtttacat taaatacact 780
 taacccctc atatttcatc tcttcctttc tcaaaaatag taataaataa cctcaagcca 840
 tta 843

<210> 102

<211> 1101
 <212> DNA
 <213> Homo sapiens

<400> 102
 gcacgagggt ctcggctcac tgcaacctcc gcctcccggc ttcaagtgat tctcctgcct 60
 cagcctcccc agtagctggg atcacgtagc gcccgcagc atgcctggct aagttttgta 120
 ttttttagtaa agacagggtt tcaccatggt ggccaggctg gtcttgaact gctgaccttg 180
 tgatccgccc gcctctgcct cctaaagtgc tgggattaca ggcatgagcc cggccgaaat 240
 acaatttaat taaaatactt attttcttat tagaaagctg cctctcaatg gcacctactg 300
 ctacatttac atagtaaccc aaaattgcag ttgcttagca gggagagaat cacagtgcgtg 360
 gatattattht atactttttc ttccaaaacg atttgaggaa gtactgtgct ggccattggt 420
 tacatcatat taggagatct ggatgtcact ttcttttccc atatcctcga tttcctcact 480
 ttttaaaatg tcattgtgtt ttgtaagttt tcttaaatcc tttggaaatg tgatgacggg 540
 gaaaaatccc tgaaggcttc tatacacctg ttggcatgga atattttgca acccgtttct 600
 tccctacaaa cagaagagac aactaaatac ggtttgatct acatctgcaa gaggctagcc 660
 attcagtatt aaaaagtgat ggccctggtt gacggtacca cacctgaaga cctatgccct 720
 ttccttcaca ctccctactt ctgcatttct tccctcctga acgtctatca agtggaccat 780
 atgaaattgc cagtattcaa ctgtttttta tcttaaaagg tgacaattct atatcattca 840
 acctaaatta atgtctcaag aacataacct ttgtttctat tattgtgacc ttacttttaa 900
 ccctcctaga gctccttaac ctgttcacac tggatttcaa ggatcttaag ttgttctact 960
 acataatcac tatcacactt cagaaacatt ttagtttaca ttaaatacac ttaaccccct 1020
 catatttcat ctcttccttt ctcaaaaata gtaataaata acctcaagcc aaaaaaaaaa 1080
 aaaaaaaaaa aaatatgcgg c 1101

<210> 103
 <211> 176
 <212> DNA
 <213> Homo sapiens

<400> 103
 gggtaacaga gtgagactcc gtctcaagag aaaaggaatt ttcttatttt aaaaataata 60
 ttctgtgtgt tatatctacc acattgtctt catttactca ttagatgtta aactgtttat 120
 tctgtatttt ggctattgtg aaaagtgtca caaacagaat tgcaaatgtt tcttca 176

<210> 104
 <211> 1689
 <212> DNA
 <213> Homo sapiens

<400> 104
 ccgctcattt tttttttttt tttttttttt tttttttttt aaacaaacaa aatttattaa 60
 actttcaaaa taaaaaaaca tcatcaaaaa gtcatagcat ttctatacat tagtaactat 120
 tcacaaatga aattcaaaaa aattccatct acctaactat agtttgaagt aaatttaacc 180
 aaaaagtgtg aacaccttac attctactct aaagaacatt atacaaatta agtaacacat 240
 aaatggaata atattactca ttcatgaact ggcattttta atagttaaatt atttgattta 300
 caaaaaatga tctgcagata taatgcaacc tctatcaaaa taccagtgcac atacttcata 360
 gacattttta aaaagcatat ctaaaattca tatggtacca caaaacaccc taaatagcca 420
 aagcaatcaa gacaaaagaa ggtatcaccg tgaactttgaa atacactaca aaactgtggt 480
 aaccaaaca gtatgtcact tgaataaaaa cagagatata ggccaatgga gcagaagaaa 540
 aagagagcag aaatatatca gtgtatttac agctaactga ttttaaatat aggtgacttt 600
 ttttttttaa gggaaataac agtatcttca ataaatgatg tttagaaaac tttatgtcca 660
 catgcagaga aaaaaaagtg agaccctcat ctcacacat atgtaaaaat aaactcaaaa 720
 taaattagcc acttaaatgt aaggcctaaa actcttaaac tactactacc aaaaaataga 780
 gtgaaagccc cataacattg gtctgggcag caagtttttt gatttaacct aaatatccca 840
 ggacacaaaa ggaagaacaa gtcagtcaga tcacttcaaa ttaaaaagct gctgcacaga 900
 atctgataca tgaacagaat gtgacaacta aaaaaaggga gaaaatattt gcaaattata 960
 catgtgacaa ggggttaata taaaaaatat atacaaaact caaatgacaa tacaacaaa 1020
 ataagtaact attaaaaata agtagctgaa ataagtattt ttcaagaaaa tacatacata 1080

36

```

atggccaact gatatatatta aaaatgctca atgtcaatta tcacaaaaag gcaagccaaa 1140
aaagaaaaaa caaaactagg agatatcaac tcattcctgt tagaatgact cttattaaaa 1200
agaaaaagcg ttggtaaaga tgtgaagaaa aggggaaggct tgcacactgt tgggttggaa 1260
tgtaaatgaa gacagccatt atgaaaaaca aaatagagat ttctcaaaaa acttaaaacta 1320
ccatgtcata acagcaattg cactactgga tatatatcca aaacaaataa aatcagaatg 1380
aagaaacatt tgcaattctg tttatagcac ttttcacaat agccaaaata cagaataaac 1440
agtttaacat ctaatgagta aatgaagaca atgtggtaga tatacacaac agaattatt 1500
ttttaaaata agaaaattcc ttttctcttg agacggagtc tcactctggt acccaggctg 1560
gatgcagtgg cacaatctca gttcactgca acctctgcat ccctggctca agggattctc 1620
atgcctcagc ctctggagta gctgggatta caggcatgca ctaccatgcc catgcccagc 1680
tgagatttt                                     1689

```

<210> 105

<211> 768

<212> DNA

<213> Homo sapiens

<400> 105

```

aaaaaattaa aagcttctag agacttctgg tttctacttc cacacataag gaacttggaa 60
attgccactc catcctatca acaagtaaaa agctaaatgg actaaaaaat caacaactct 120
tataagacgg aaagtcactg agtatgatgc tgcctcccaa cttggagaat acagggagtc 180
acatctctcc agagtggaga ttcatgagaa gaaacaccaa tgagaaaaag aaatggagta 240
tgaaacctga actctaattg atgaatttct ggagaataag tgaggacaag actgagaatt 300
aaacattcca gaaaaactaa ctcataaggg gaacttcaca atattttgag attcaccttc 360
acaaatttga ccattttcca cagcaaatat cagagaaaaa ttaacttgta cattcaggag 420
agaaagggaa aaagaaacct ctttgaata taccacagag ctctattcct cttatcaagg 480
cctgccctca gaagaaacga attaaccaaa actatcatca gagcctaatt gacctgggga 540
agagaaatgc ttgtctctg ctccactagt tttctacctg tgagaaggca aatacacaac 600
tccagccac tctagtcatc ttgtcctacc aaagcgggag aacaaaacag aacaacactt 660
gtaaagtga caatccagac gcatagactc actaaaagc tgagatgtaa tcattaaact 720
aaaatccttc ccctgccact acaccatatt actaaaggcc tatttaga 768

```

<210> 106

<211> 612

<212> DNA

<213> Homo sapiens

<400> 106

```

gggaatttca gacaacctag cctagactaa atgggtgggca gcacctggca gacaagaact 60
caagaacctt ttctcagggt gctctgcttt gctgcaggta atggagaagc actggagatt 120
tgtaagccac ggagtcaaat ggtggactgg gatcttcagg agatcattta gagagcaaga 180
tcttaccaaa tcttttagtc atggcttatt tcgttgcaat catatgggtt ttactgcaaa 240
ggtgaagaac taatgactgc agcaggaaaa agaattggat gtgtcatgaa ttatggccct 300
gcttatactt ctacttcaac cgtaatcatt tgtttaaaca aaaagttctg catttgaatt 360
gtcacaaattg tgtgtgtgtt ataaacatct catatttcat ccaggctcag ccaacacttg 420
cctttattaa tgcataat caagaaataa atctcactat aaccaaaaat tatccttcat 480
aagagaatat aaacagaagt ctggttcata aacttactaa ttaacacctc tattctcatg 540
tatcaactaa catttttggt tcgtcttaaa ataaataaaa ctttatgaca tgctaataat 600
ttatttaaaa aa                                     612

```

<210> 107

<211> 628

<212> DNA

<213> Homo sapiens

<400> 107

```

aaattatttg caaacacttt ttagctgaac cctctcattt cacagtggag ccttttaattg 60
tttcctttgc agaactcaag aaccttttct cagggtggctc tgctttgctg caggtaattg 120

```

37

```

agaagcactg gagatttcta agccacggag tcaaattggtg gactgggatt ttcaggagat 180
catttagaga gcaagatctt accaaatcct ttagtcatgg totatttcgt tgcactcata 240
tgggtgttac tgcgaagggtg aagaactaat gactgcagca ggaaaaagaa ttggatgtgt 300
catgaattat ggccctgctt atactttctac ttcaaccgta atcatttgtt taaacaaaaa 360
gttctgcatt tgaattgtca caattgtgtg tgtgttataa acatctcata tttcatccag 420
gctcagccaa cacttgccctt tattaatgct cataatcaag aaataaatct cataactaacc 480
aaaaattatc cttcataaga gaatataaac agaagtctgg ttcataaact tactaattaa 540
cacctctatt ctcatgtatc aactaacatt tttgtttcgt cttaaaataa ataaaacttt 600
atgacatgct aataatttat ttaaaaaa 628

```

<210> 108

<211> 103

<212> DNA

<213> Homo sapiens

<400> 108

```

ctagaccacg ttgtggaaat gtctcacaac attgatctac taggcaagga tttttgaggt 60
cagaccgcaa aaaccacagg gcaaccaaaag gccaaagtta gac 103

```

<210> 109

<211> 348

<212> DNA

<213> Homo sapiens

<400> 109

```

gtgaatcctt gtaatcctcc gtctccagac ggcagtggcc agagtggacg tgggtggcctg 60
agctgtggcc tgggctgtgt ctggaggctg ggatttgggc tccggctctg tcccagccca 120
gatgctggtc ccttccactc tggtcaggtc agtgaataga gcaccagga aatgggtgct 180
gcggtcatag ttgtggctgt gggttattaat aacactgtcg tgttactgtt atgagagagt 240
gtggtgagag catctgtccc agcctagcag gccacagact ttctagaggg gcagtagagg 300
tagaaacaac tcaggattct gagagtcctc aagtccatcc tggccctg 348

```

<210> 110

<211> 616

<212> DNA

<213> Homo sapiens

<400> 110

```

cgaggctggc ggtgcgctgc ttctcagag ccgcttctc agagccggct gcggcgggcc 60
cgggcgggaa ccacggagcc cagtgcacca gcctcctcgg tgctaccgcy ggacacagag 120
gaaacaggaa cagctgggtt ctgtgggcag gccccgggct ggaactagag ccagggtgcy 180
gccggcgggg gacagggaaa gagatcacag cgaagacca gaagaaacaa aaggcaagcy 240
aatattttta tatccaactg cctactggac accaaccacg tggacaagtc ctggttgcc 300
caaactcaac atgttcaaag ctgaatacat cacctgctct cccaaatatg ctctctctct 360
gctgttccca aaatcagaaa atggcttcac gatcagctca gtcattctca gagcaaatgc 420
tgagagtcac ccttgaatcc ttctgttgcc tccacattca aaccatcacc atatccttga 480
tttctctact gtatattttt catatgtgtc caattcttct catctgcaat ctcattagtg 540
aaggccacca acatctctca tctgaatgcc tgcaatacct cctcacaggc caccaggcat 600
ctagttttgc ccctgt 616

```

<210> 111

<211> 1049

<212> DNA

<213> Homo sapiens

<400> 111

```

atgagctccc gagcttgggt tctgaagtg gattatgctg gaggaacaca ggtagaagca 60

```

38

```

gaagtaacaa aggagagaag gagactgccc tactgccccta taccaggaag gaataaagcc 120
aaaaaaacag aattctccaa gtgtcaagca aaaacacata ctttgacac gtttctcgag 180
gtccagcccc aaagcctgcg ccctggggcg tccctgcttc ggccccaga ggggggcagg 240
cctcgctcct ccctccgcca ggcctgcccg ggaggcctcg acccgcgag gtgaccgccc 300
ccagggctcg cggcgcgagg acgaggctgg cgggtgcgtg ctctctcaga gccgcttcct 360
cagagccggc tgcggcgggc ccgggcggga accacggagc ccagtgcacc agcctcctcg 420
gtgctaccgc gggacacaga ggaaacagga acagctggtt tctgtgggca ggccccgggc 480
tggaactaga gccagggtgc ggccggcggg ggacaggga agagatcaca gcgaagacc 540
agaagaaaca aaaggcaagc gaatatTTTT ataccact gcctactgga caccaaccac 600
gtggacaagt cctggttgcc tcaactcaa catgttcaaa gctgaatata tcacctgctc 660
tcccaaatat gctcctctcc tgctgttccc aaaatcagaa aatggcttca cgatcagctc 720
agtcattctca agagcaaatg ctgagagtca ccctgaaac cttctgttgc ctccacattc 780
aaaccatcac catatccttg atttctctac tgtatatTTT tcatatgtgt ccacttcttt 840
ccatctgcac tctcattagt gaaggccacc aacatctctc atctgaatgc ctgcaatacc 900
tcctcacagg tcaccaggca tctagttttg cccctgtcct gcccttccct catctagagt 960
gaagccagta ggaaccttcc aaaatgaaaa tctgattaag tcacttcttt gcttaaaact 1020
tttttatggt ttcacagccc atgaaaata                                     1049

```

<210> 112
 <211> 388
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (324) .. (364)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (364)
 <223> a, c, g or t

```

<400> 112
gtgaccttgc actcccctgg cctgaagctg cctctctgcg cgctttctac tgggctcgtc 60
tctttccgga gccccagcgt ctctgccc aattcaccgc ggaaagggcc cggggcgagg 120
gtgcgaccgg gcgtcggcag cgcagacctc ttggccttct ctccaggtc ggtgcgctcg 180
ctctccgctg tccccgcccg actgcccgtg agtccatggc tagacgcgcc ggacaggact 240
gatggcggga ccgcgctgcc cgagaaaggg acggaccaat acgtgtgttt cctccgctat 300
cagtcctgct ccttcgggca cctncggggc ccggcggtg gctaattgtt tgtttgaaag 360
atcngtggaa tttttaagag agtatTTA                                     388

```

<210> 113
 <211> 756
 <212> DNA
 <213> Homo sapiens

```

<400> 113
gcggcgcgcg caccgcgcgc tgccccaccg caccacgggg ccgcccgcgc gccgcggggc 60
cagctcagcc ctgccagccc agccaccgcc gcgccccggg cgcccgcgct gcattcgcg 120
ctcgatctct gagagcccac cgcagccgg tgacagcga tgccaggatg caggagcgcg 180
cgacgcggcg ccgggtcgca gccgacgacg ccgcccgcag cctgacctca caccctctgg 240
gccccgcctc tggagccagc gccagggttc cctctgtgct ttttcgcttt cctaagctcc 300
tgtcgctcct ctttgtcccc tcagtttatg tcctcctgtg ctccacctcc tgacctctgt 360
gaccttgcac tccccgggcc tgaagctgcc tctctgcgcg ctttctactg ggctcgtctc 420
tttcgggagc ccagcgtct cctgccccaa ttaccgcgg aaagggcccg gggcgagggt 480
gcgaccgggc gtcggcagcg cagacctctt ggccttctct cacaggtegg tgcgctcgct 540
ctccgcgttc ccgcccgcac tgccgtgcag tccatggcta gacgcgcgg acaggactga 600
tggcggggacc gcgctgcccg agaaagggac ggaccaatac gtgtgtttcc tccgctatca 660

```

39

gtcccgtcgc ttcgggcacc tccgggcccc ggcggctggc taatgttttg tttgaaagat 720
 cgggtggaact ttttaagaga gtatttaaaa aaaaaa 756

<210> 114
 <211> 918
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (314)..(342)
 <223> a, c, g or t

<400> 114
 cgcgcgggac aggactgatg gcgggaccgc gctgcccag agaggacgg accaatacgt 60
 gtgtttgctc cgcgaaccct cttgaagctg ttcagaagcc gcttgccgcg gggccacta 120
 ggcggggcgg gggttgggac ccagcgggag ccggggcagc ctggctccac ggccgtact 180
 cggtttacac cgcgggcggg cgcggaggga ggctgcgttt cctccgctat cagtcccgtc 240
 gcttcgggca cctccgggcc ccggcggctg gctaatgttt tgtttgaaag atcgggtgaa 300
 ctttttaaga gagnnnnnnn nnnnnnnnnn nnnnnnnnnn nnttcaccgg gcaaccgggg 360
 aagtattgtg gccttgaggt ttgctaaatc caaatatgaa aatcaaaagc ttagtattc 420
 ctcatcttct cttctggaag atttgcgtta gaggttttgt tgggccttca aaaagctgtg 480
 ttcagagtta ggagaatata tccaataaaa gatggtttcg tctaccaatt ggggaagttt 540
 caccctctcc ctatctgaag aaaaaaatca aaaacaaatg tcccggatc ttcgatgca 600
 agtcctggag gcaggagat cactgcctgc ctggcccacg ctgctgggac ggctcgctcct 660
 ccctgctttt tgtttttcaa acctcctgct tctcccacct tgggaaggag aaatgtgaaa 720
 cccggcagcg gccgacctag gcggtcttgt ggcgggagc cggccgggcc cgaaaacct 780
 agacctggtt gtactgtagc ttgttgtttg ggggaccaa ttttctagag agaactagag 840
 cacttttgtt gtgttttttt gttttgtttt tgttttttgc cttgtcgatt cccgaataaa 900
 ttttgtgttc cttctttt 918

<210> 115
 <211> 2753
 <212> DNA
 <213> Homo sapiens

<400> 115
 tgggcgggact ccccatggcc agaggctgag ctccactccc gccggccgct ccctagggga 60
 aggggaaagga gaggggagag cagcgacagc cctccagcaa gcaaagcgcg ggcggcatcc 120
 gcagctctca gaagtttgag acttgccgt aagcggactc gtgcgccccca actctttgcc 180
 gcgccagcgc ctggagcggg gagcagaggg ggcccgcccg cggcgccgcg gctttgtcat 240
 gatggccagc taccctgagc ccgaggagcg ggccggggcc ctgctggccc cagagaccgg 300
 tcgcacagtc aaggagccag aaggggccgc gccgagccca ggcaagggcg gtgggggtgg 360
 cggcgggaca gcccgggaga agccggaccc ggcgcagaag ccccgtact cgtacgtggc 420
 gctcatcgcc atggcgatcc gcgagagcgc ggagaagagg ctacagctgt ccggcatcta 480
 ccagtacatc atcgcaagt tcccgttcta cgagaagaat aagaagggtt ggcaaaatag 540
 catccgccac aacctcagcc tcaacgagtg cttcatcaag gtgccgcgcg agggcggcgg 600
 cgagcgcaag ggcaactact ggacgctgga cccggcctgc gaagacatgt tcgagaaggg 660
 caactaccgg cgccgcccgc gcatgaagag gcccttccgg ccgcgcgccg cgcacttcca 720
 gcccggaag gggctcttcg gggccggagg cgccgcaggg ggtgcgggc tgggcgggcg 780
 cggggccgac ggctacggct acctggcgcc cccaagtac ctgcagctct gcttcctcaa 840
 caactcgtgg ccgctaccgc agcctccctc acccatgccc tatgcctcct gccagatggc 900
 ggcagccgca gcggctgcag cagctgcggc tgcagccgcg ggcggcggtg gccctggcgc 960
 ggccgctgtg gtcaaggggc tggcggggcc ggccgcctcg tacggggcgt acacacgcgt 1020
 gcagagcatg gcgctgcccc ccggcgtagt gaactcgtac aatggcctgg gagggccgcc 1080
 ggccgcaccc ccgcctccgc cgcaccccca ccgcacatcc caccgacacc atctgcacgc 1140
 ggccgcggca ccgcgcctg cccacccgca ccacggggcc gccgcgcgc cgccgggcca 1200
 gctcagccct gccagcccag ccaccgccc gcccggggcg cccgcgcca ccagtgcgc 1260
 gggcctgcag ttcgcttgtg cccggcagcc cgagctcgcc atgatgcatt gctcttactg 1320

40

```

ggaccacgac agcaagaccg gcgcgctgca ttgcgcctc gatctctgag agcccaccgc 1380
atgccgggtgc atgacggatg cgaggatgca gggacgcgcg acgccggccc cggctcgagc 1440
cgacgacgcc gccgccagcc tgacctcaca ccctctgggc cgcctctgag agccagcgcc 1500
cagggtccct ctgtgctttt tgcctttcct aagctcctgt cgctcctctt tgtcccctca 1560
gtttatgtcc tcctgtgctc acctccctga cctctgtgac cttgcactcc cctggcctga 1620
agctgcctct ctgcgcgctt tctaactgggc tcgtctcttt cgggagcccc agcgtctcct 1680
gccc aaattc accgcgga aa gggcccggg cggaggtgcg accgggcgtc ggcagcgag 1740
acctctgggc cttctctcac aggtcgggtg gctcgtctc cgcgttcccc gcccgactgc 1800
cgtgcagtcc atggctagac gcgcgggaca ggactgatgg cgggaccgcg ctgcccgaga 1860
aagggaacgga ccaatacgtg tgtttgctcc gcgaacctc ttgaagctgt tcagaagccg 1920
cttgccgcgg gccccactag gcggggcggg ggttgggacc cagcgggagc cggggcagcc 1980
tggtccacg gcctgtactc ggtttacacc gcggggcggg gcggaggag gctgcgtttc 2040
ctccgctaat agtcccgtcg cttcgggcac ctccgggccc cggcggctgg ctaatgtttt 2100
gtttgaaaga tcggtggaac tttttaagag agtattttaa aaaaaaaaaa aaaaaaaaaa 2160
ttcacggggc aaccgggga gtattgtggc cttggagttt gctaaatcca aatatgaaa 2220
tcaaaagctt tagtattcct catcttctct tctggaagat ttgcgttaga gtttttgtt 2280
ggccttcaaa aagctgtgtt cagagttagg agaatatat caataaaaga tggtttcgtc 2340
taccattgg ggaagtttca ccctctccct atctgaagaa aaaaatcaaa aacaaatgtc 2400
ccggatcct tcgatgcaag tcctggaggc agggagatca ctgcctgcct ggcccacgct 2460
ggtgggacg ctcgtcctcc ctgctttttg tttttcaaac ctctgcttc tcccaccttg 2520
ggaaggagaa atgtgaaacc cggcagcggc cgacctaggc ggtcttgttg cccggagccg 2580
gcccggcccg aaaaccatag acctggttgt actgtagctt gttgtttggg ggaccaaatt 2640
ttctagagag aactagagca cttttgttgt gtttttttgt tttgtttttg ttttttgcct 2700
tgtcgattcc cgaataaatt ttgtgttctt tcttttaaaa aaaaaaaaaa agg 2753

```

<210> 116

<211> 81

<212> DNA

<213> Homo sapiens

<400> 116

```

gttgcaatat tttctcttc ctgttttgac cttgctcatg gtgcctttta tttttattta 60
attaattaat ttattgtcta a 81

```

<210> 117

<211> 558

<212> DNA

<213> Homo sapiens

<400> 117

```

gaaagtaagt taagaagagg aaatcaaagt gagctgtcta atctttaagt aggcattaca 60
ataacaattg attagttctg ccaattcttt tacaattttg gttatctaca ctttatttct 120
gtgtgtataa gtggaatcac aggcctgctt tactgctgtg atgcagtagc ttgaattgtg 180
ctataaatag catattttgc ctgtaatatc aactataagc attctctata atcaagcaat 240
tatgcctcta aagcacataa aatttaaaaa tctgttctta ttagctctgg aaatattgtg 300
gaattttaca tggaatctta tcttggaag gtagattttg aaattcttag aggattattt 360
gtccccattt ccattcagct gacatggtga cttttgtcac aagtcctaaa aattagaata 420
atcagagggc aagggggaca tcaactgcag atgttgagga agcctagtgc aatttagaat 480
aaattttact atttaaaact cacctattgc tcagagagca attatatatt ggtaggaatg 540
actcatctat gggctaaa 558

```

<210> 118

<211> 693

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (209)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (444)
 <223> a, c, g or t

<400> 118
 gtcacacaca ctcttctgac actgacgacc tggagtgtca cagaccctga aggtgaaggg 60
 ctctgtccca caagactctg cccctacttc tgatgacagc cgtacatggg taccaggga 120
 acccacactc actcctgaca actgcagatt tggggaactt tacatcccct cagattcact 180
 agaacacctc ccagggtca ggaagtgtt ttacgtacaa tcatgcttat tatgaaggaa 240
 acccatgaac agctcagtga agagagtggg gaggtgggca tgatctctga gcaccgtggg 300
 ggctccccag cctgggggct ccccaaccct gatgcccata agtttttctc taggcctcat 360
 tacacaggta tgattgatta agtcattggg cattggtgat tgaacacaaa ctcaatctct 420
 ggccccctcc aggagtggg gcgntgagg gggttggaag ttctctctta attacatggt 480
 tgggttctct ggcaacaagc tcccacccta aagctacctt ggggtcccc aagagtcacc 540
 tcattagggt aaacaaatgt ggtgaaaaag agttgttatg aaatcagaca cccctatcag 600
 gaaattccaa agatttaagg agttctgtcc ctggaacagg ggacaaagac cagatgtatt 660
 ttttattata ccacaatata aatctcttaa ttt 693

<210> 119
 <211> 838
 <212> DNA
 <213> Homo sapiens

<400> 119
 tcacacacac tcttctgata ctgacgacct ggagtgtcac agaccctgaa ggtgaagggc 60
 tctgtccca aagactctgc ccctacttct gatgacagcc ggtacatggg taccaggga 120
 acccacactc actcctgaca actgcagatt tggggaactt tacatcccct cagattcact 180
 agaacacctc ccagggtca ggaagtgtt ttacgtacaa tcatgcttat tatgaaggaa 240
 acccatgaac agctcagtga agagagtggg gaggtgggc acctgatctc tgagcaccgt 300
 gggggctccc cagcctggg gctcccaac cctgatgccc aaaagttttt atctaggcct 360
 cattacacag gtatgattga ttaagtcatt ggtcattggg gattgaacac aaactcaatc 420
 tctggcccc cccaggagtg gggcggtga gggggctgg aagttcctct ctaattacat 480
 gggttggtcc tctggcaaca agctcccacc cttaaagctac cttgggggtcc cccaagagtc 540
 acctcattag ggtaaacaaa tgtggtgaaa aagagttgtt atgaaatcag acaccctat 600
 caggaaattc caaagattta aggagttctg tccttggaac aggggacaaa gaccagatgt 660
 attttttatt ataccacaga agagtaataa gacgaacata tataccagc atccaaatta 720
 agaaacataa cataaaggta tcttttaagc ctcttggtgt cctttgtgaa tatatttct 780
 ctgcttccca gaggaacca ttatcttgaa ttttgtgtta tctgttacct tgcttctc 838

<210> 120
 <211> 551
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (494)
 <223> a, c, g or t

<400> 120
 gtaacttcct taacatcaca ttgcttggag atatcagttt ggctgttcat ttctaattta 60
 gattgtttcc aaatgttcag aattaaaatc tgtatactta aattctgtac atagatcact 120
 ttgggagttc tgaaatatc atgaatactt gcacctttt ccagaatcta aacttcatac 180
 atctagtttt gttcttgtaa attgttttga ggaagtggg gtcagtgtca caaacagct 240
 gtggctccaa acagacacca ggatttaggc ccattacaga gagaccacc tggaaatatt 300

42

```

ctacagttga gaggagcttt cagtctagaa gaggaggaaa tgatacttag tttagtcata 360
atgtgctttg gcaagaaatt acagtcgaaa ggaaggaaca gataaacatt gtgtggtgta 420
gccactttga agagtgggtca aattccctgt ggcaaaactt cctcctcccc tcttcattcc 480
ccattccccc tatnttgatg ttagataggt ggcactttac tgtgtcactc ccggcctatn 540
ctccccacaa c 551

```

```

<210> 121
<211> 635
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (540)
<223> a, c, g or t

```

```

<400> 121
gtaacttctt taacatcaca ttgcttggag atatcagttt ggctgttcat ttctaattta 60
gattgtttcc aaatgttcag aattaaaatc tgtataacta aattctgtac atagatcact 120
ttgggaggtc tgaaaatatt atgaataact gcaccttttt ccagaatcta aacttcatac 180
atctagtttt gttcttgtaa attgttttga ggaagtgggt gtcagtgtca caaaccagct 240
gtggctccaa acagacacca ggatttaggc ccattacaga gagaccaccc tggaaatatt 300
ctacagttga gaggagcttt cagtctagaa gaggaggaaa tgatacttag tttagtcata 360
atgtgctttg gcaagaaatt acagtcgaaa ggaaggaaca gataaacatt gtgtggtgta 420
gccactttga agagtgggtca aattccctgt ggcaaaactt cctcctcccc tcttcattcc 480
ccattccccc tattttgatg ttagataggt ggcactttac tgtgtcactc ccggcctatn 540
ctccccacaa cactacttgg agtttaatca taagatcgtg gttttatttt tttcccttaa 600
aagatggatc tttattttct ttacttttat attct 635

```

```

<210> 122
<211> 118
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (99)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (113)
<223> a, c, g or t

```

```

<400> 122
attcagggct ctttccattt taccacacta ttcaaaattt ggattctcta tgtagccaaa 60
tggataatga gaacccaaaac ataaaaaaag agaagaaana aaaaaagaaa ganaaaga 118

```

```

<210> 123
<211> 673
<212> DNA
<213> Homo sapiens

```

```

<400> 123
tttttttttt ttgagacaga gtctcgtctt gtcgccagg ctggagtga gtggtgcaat 60
ctcagctcac tgcaaacctc cgctcccggt gttcatgcca ttttctgccc tcagcctccc 120
aagtagctga gactacgggc acacgccaca acgcccgggt aattttttgt atttttagta 180
gagacagggt ttcaccgtgt tagccaggat ggtctcgatc tcctgacctc gtgatctgcc 240

```

43

```

tgccctcggt tcccaaagtg ctgggattac aggcgtgagc caccgagccc agcctaaaaa 300
ctatttttat atattctctt tacatctcca taatcctgta aggacgtagg cattattctt 360
tttttctaga taattgccat aataaattca tggaatcagt gtagggaaga caaaaaaaga 420
aaaaaaaaat tcagatgaga aaactaaggg acttgctcaa agctgcacaa ctagtaggaa 480
cagaataaacc caattcttac agtgtcttca ttcagggtc cttccatttt accacactat 540
tcaaaatttg gattctctat gtagccaaat ggataatgag aacatgtata aaataataaa 600
gaaataaaact acaatcataa aaagtaacta aaatagccaa ctgtcatgta aaaggtatgt 660
agcaaaactga cag                                     673

```

<210> 124
 <211> 370
 <212> DNA
 <213> Homo sapiens

```

<400> 124
ggggagagca gagcagagcg tgaagggtgct gggaggcctg cctcaaagtt ggcaaaaccc 60
acagcgtctc agagctgctg tcatgttcta gttcctgcct ctgtgccagt gagaccagaa 120
aaccaggcca ctcaaaagcc tcttgctgtg gctctctatg aatggagggt ggggcaaggg 180
caggaccctt gggcctcagg cgagaagaag cagatttacc ctgagctttc ttctgtcttg 240
tggcattggc tgtgccccgg attttaggag ccttgccctt tctcatccga gaagcacctc 300
taacgcgaac cctccttcgc gcantatagc tgcaaatgat aaccgtcttt gaattgtaca 360
aaagcttatg                                     370

```

<210> 125
 <211> 896
 <212> DNA
 <213> Homo sapiens

```

<400> 125
cacaagacat agcagcagag gtgcacagcg ctcagcagtg acctcgcatg caccgagggt 60
ggacccccga accaaatctc cctgggtcca cttttcacia gctgcgcgaa ggggacaagt 120
cctgccacct ataagcctcc gtttccatgt ctatacactg gggttcctag ctcacgggac 180
tgtcggggta attgagttag ttaacgtcta gggagcacct gtgacatgcc aacacagtgc 240
tgtcatttct gctgttgctc atttttctgc atctttatct gtaaggattt gaaagaatgt 300
acagttggaa acctgatgat ctcaagcaga aaatatcttt tcataacgct gagcatgaat 360
gacatgagaa tccatgtctg aagtgaatc gtatggatct gaagaatggg tgggtgccagc 420
cctggtggaa tggggtgcga aggagggagg atgagagcca gacgtttcag tctgggtgac 480
cctgccaccc agagccacct tccattaaact gaggggtcca gggctccctc cggggcactt 540
gccactaaag ctcagctaaa gtctcaaaaa ggacacattc ggagccaagc aacaggcaca 600
gcccattgta ggaatgtttc tgcaatggaa aaatacaaaa ccagaaaagg agtgtgtggg 660
cctaactcga catgtttatc aacattttac tgcaatgtat gacattttct tgagcacaag 720
attagccttg gtattttttt ctgggaagta taaaagactt tttttttctt tcttttggtt 780
ttcaatttct ctctagagga atttaaaacc ggatatttcc atcttaaggt tcttgagcaa 840
gtctgtcagg gtgtccatat ttcttaccct gttcctctca gcatcgaagt gctatc 896

```

<210> 126
 <211> 998
 <212> DNA
 <213> Homo sapiens

```

<400> 126
cacaagacat agcagcagag gtgcacagcg ctcagcagtg acctcgcatg caccgagggt 60
ggacccccga accaaatctc cctgggtcca cttttcacia gctgcgcgaa ggggacaagt 120
cctgccacct ataagcctcc gtttccatgt ctatacactg gggttcctag ctcacgggac 180
tgtcggggta attgagttag ttaacgtcta gggagcacct gtgacatgcc aacacagtgc 240
tgtcatttct gctgttgctc atttttctgc atctttatct gtaaggattt gaaagaatgt 300
acagttggaa acctgatgat ctcaagcaga aaatatcttt tcataacgct gagcatgaat 360
gacatgagaa tccatgtctg aagtgaatc gtatggatct gaagaatggg tgggtgccagc 420

```

```

cctgggtggaa tgggggtgcga aggagggagg atgagagcca gacgtttcag tctgggtgac 480
cctgccaccc agagccacct tccattaact gaggggtcca gggctccctc caggccactt 540
gccactaaag ctcagctaaa gtctcaaaaa ggacacattc ggagccaagc aacaggcaca 600
gcccattgta ggaatgtttc tgcaatggaa aaatacaaaa ccagaaagga agtgtgtggg 660
cctaactgta catgtttatc aacattttac tgcaatgtat gacatttctg tgagcacaag 720
attagccttg gtattttttt ctgggaagta taaaagactt tttttttctt tctttttgtt 780
ttcaatttct ctctagagga atttaaaacc ggatatttcc atcttaaagt tcttgagcaa 840
gtctgtcaaag gtgtccatat ttcttaccct gttcctctca gcacgaagt gctatctctg 900
ttacactcat gtttgctgtt cacaatggag tactaatgaa atagcaaat taagctaccg 960
gcatggtgct aataactgaa actaaaaatc ggttgag 998

```

```

<210> 127
<211> 838
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (100)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (112)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (134)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (151)..(152)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (161)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (164)
<223> a, c, g or t

```

```

<400> 127
agggcataaa cacttttagt tgatcagtag aattgctatg ccatgtttta atgggattta 60
tttggttgat gcagaatata taattgtatc tagaagatan atattacaaa antattttta 120
tatacaattt ctgncatatt tttgggaaag nncatttttg ngngncaaag tagaatcatt 180
gttgccaata gagttagcat ctttgtgtgc ttgtgagggt tgattttgag ggttttcttg 240
gttttgttt gggttctgga gttctaaaa atgagattgt ctttgtctaa acaattttta 300
tataaaaaatg tacatttttg tattattttt tcttattcca acctaatcgg tggcttggtc 360
cttctgtgtg ttattgggct gttgggtgcc tggatagagc tggagaccat ttaactgctg 420
tatgaataat agataagcgt cttgaataac atctgaattt cctaggtatg tagaaacacc 480
caccatgcac atatatgaac atacagaata tatgaatgtt aaaatatggt gaaaacaatc 540
ttttgctaata agaagtgtta acctttattt ttaaaaaaaa tttggtgtgt atgtagaggt 600
ttatttgatt gttagtgtg tccatgtata atatgtcatc tacctttaca gatgtgcaga 660
aatttggtgt atttggtgga tatattttac ttaaaactat agggcagaag ctttttatgt 720
ttgttgaagt gaaatggcat accaaacctg tgtggtagag tgggattttt agattgctgt 780

```

gtgtacagtc aggttatatc tttaaaatac ctattcggtta tatattaata tgtagaca 838

<210> 128
 <211> 5542
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (5379)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (5382)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (5391)..(5392)
 <223> a, c, g or t

<400> 128
 cacaaacccg gaagcggatc gcgtggagtg aaggtcctac cacggcgcgt gagtttcgct 60
 ctgccttggg ttaagtctgc acttcccagg tccccggcgc ttctgcccct gggacgtggg 120
 atccccacgg acctggaaat tctcgccgtg cttcccttca cccagagcaa attgagacgt 180
 cccggaggaa gaccaaggca gcctattggg ccttccaggc aatcacatgg gaatcagcca 240
 cacgtcattc ctctcacct cagaacatct cagaataact tggtgaaatg tctcccactg 300
 tgagcctcag tgagcccacc tgtaacatag aggcctcgcc cctgagctct acaatcctgt 360
 gtccagttgt ctctcagct gtctcctggg tcatcaaagc ggcacccca ccttcagggtg 420
 tccacgagtg gctttctaaa cccccaaca catttccttg cagtctgcac atctcagatg 480
 aggggtgacta cgtacttccg gaaacggccg aacttgacag catgtatttt aaatttgtga 540
 aataaattac tttatttga agtggtgtaa ttataatat aaagagaaac ttagatgtat 600
 acgtgaaaag agtgagaaga tacatcactt ccaattttgt ttgtttgttt gtttttttga 660
 gaggaatttt cactcttggt gctgaggctg gagtgcattg ccatgatata agctcactgc 720
 aacctctgac tcttgggac aagggattct ccttctcag actcccgagt agctgggac 780
 acagtcgact ttcaaaattc tttaaggatt gattcctaaa gactcatgtt atgtgaagaa 840
 gcagctcaga agaggaaaag aaaggagcca ggcattggctc ttctcagggt acgcttgact 900
 ttcagggtatg tggctataga attctcattg gcagagtggg aatgcctgaa ccttctgcag 960
 agggctttgt acagggaagt gatgttggag aactacagga acctggaagc tgtggatatc 1020
 tcttccaaac gcatgatgaa ggaggtcttg tcaacagggc aaggcaatac agaagtatc 1080
 cacacaggga cattgcaaag atatcaaagt tatcacattg gagatttttg cttccaggaa 1140
 attgagaaaag aaattcatga tattgagttt cagtgtcaag aagatgaaag aaatggccat 1200
 gaagcaccga tgacaaaaat aaaaaagttg actggtagca cagaccaaca tgatcacagg 1260
 catgctggaa acaagcctat taaagatcag cttggatcaa gcttttattc acatctgcct 1320
 gaactccaca taattcagat caaaggtaaa attggtaatc aatttgagaa gtctaccagt 1380
 gatgctccct cggtttcaac atcccaaaga atttctccta ggcccaaat ccatatttct 1440
 aataactatg ggaataattc cccgaattct tctactactc cacaaaaaca ggaagtatac 1500
 atgagagaaa aatctttcca atgtaatgag agtggcaaag cctttaattg tagctcactc 1560
 ttaaggaaac accagatacc ccatttagga gacaaacaat ataatgtga tgtatgtggc 1620
 aagctcttta atcacaagca ataccttaca tgccatcgta gatgtcacac tggagagaaa 1680
 ccttacaagt gtaatgagtg tggaaagtcc ttcagtcagg tatcatccct tacatgccat 1740
 cgtagacttc acactgcagt aaaatctcac aagtgtaatg agtgtggcaa gatctttggg 1800
 caaaattcag cccttgaat tcataaggca attcatactg gaaaaaaacc ttacaagtgt 1860
 aatgaatgtg acaaagcttt taatcagcaa tcaaaccttg cacgtcatcg tagaattcat 1920
 actggagaga aaccttaca atgtgaagaa tgtgacaaag ttttcagtcg gaaatcaacc 1980
 cttgagtcac ataagagaat tcatactgga gagaaccat acaaatgtaa ggtttgtgac 2040
 acagctttca catggaattc tcagctggca agacataaaa gaattcacac tggagagaaa 2100
 acttacaagt gtaatgagtg tggcaagacc ttcagtcaca agtcatccct tgtatgccat 2160
 catagacttc atgggtggaga gaaatcttac aaatgtaagg tctgtgacaa ggcttttgcg 2220

tggaattcac	acctggtaag	acatactaga	attcatagtg	gaggaaaacc	ttacaagtgt	2280
aatgaatgtg	ggaagacctt	tggtcaaaat	tcagatcttc	taattcataa	gtcaattcat	2340
actggagagc	aaccttacaa	atatgaagaa	tgtgaaaagg	ttttcagttg	tggatcaacc	2400
cttgagacac	ataagataat	tcacaccgga	gagaaacat	acaaatgtaa	ggtttgtgac	2460
aaggcttttg	cgtgtcattc	ctatctggca	aaacatacta	gaattcatag	tggagagaaa	2520
ccttacaagt	gtaatgagtg	cagcaagacc	ttccgtctga	ggtcatacct	tgcaagccat	2580
cgcagagttc	atagtgggtga	gaaaccttac	aagtgtaatg	agtcgagcaa	gaccttcagt	2640
cagaggttcac	accttcattg	ccatcgtaga	cttcacatgt	gtgagaaacc	ttacaagtgt	2700
aatgagtgtg	gcaagacctt	cagtcacaag	ccatcccttg	ttcaccatcg	tagacttcat	2760
actggagaga	aatcttacaa	atgtacgggt	tgtgacaagg	ctttcgtgcg	taattcatac	2820
ctggcaagac	ataccagaat	tcacactgca	gagaaacctt	acaagtgtaa	tgaatgtggg	2880
aaggctttta	atcaacaatc	acaactttca	cttcacata	gaattcatgc	tggggagaaa	2940
ctttacaaat	gtgaaacatg	tgacaaagtt	ttcagtcgca	aatcacacct	taaaagacat	3000
aggagaattc	atcctggaaa	gaaaccatac	aaatgtaagg	tttgtgacaa	gacttttggg	3060
agtgattcac	acctgaaaca	acatactgga	cttcacactg	gagagaaacc	ttacaagtgt	3120
aatgagtgtg	gcaagacctt	tagcaagcag	tcaacactta	ttcaccatca	ggcagttcat	3180
ggtgtaggga	aacttgacta	atgtaatgat	tgtcacaaag	tcttcagtaa	cgctacaacc	3240
attgcaaatc	attggagaat	ctataatgaa	taaagatcta	acaagtgtaa	taaatgtggc	3300
aaatttttca	gacatcattc	atacattgca	gttcattgac	acactcatac	tggagagaaa	3360
ccttacaaat	gtcatgactg	tggcaagggtc	ttcagtcagg	cttcaccta	tgcaaaacat	3420
aggagaattc	atacaggaga	gaaacctcac	atgtgtgatg	attgtggcaa	agcctttact	3480
tcattgtcac	acctcattag	acatcagaga	atccctactg	gacagaaatc	ttacaaatgt	3540
cagaagtgtg	gcaaggtctt	gagtcaggag	tcactccttg	cagaacatca	gaaaattcat	3600
ttttgagata	actgttccca	atgcagttag	tatagcaaac	catcaagcat	taattgacac	3660
tagagttagt	tcagcattga	cttgagtttg	acttaacatt	gagttgaagc	cttaattgac	3720
attaaagtgt	ttatgttaag	aggactgggc	caggcacagt	ggctcacacc	tgtaatctga	3780
gagctttggg	aggccagcac	cggtagatca	cttgaactcc	cagcctcaga	tgatccaccc	3840
acctcgacct	cccaaagtgc	tgggattaca	ggcgtgagcc	tccgcacccg	gccaaagatat	3900
accattcaat	gtagagatta	ttcttacatg	aaactgacct	aaacaattta	ataaaaatta	3960
atttttactt	ttaaatcaaa	atgtggggga	gggaacctact	ctccctctct	aacatgacag	4020
catatatata	tttaatatat	aagcttaaat	atgtgcaagt	aatttgtctt	ttacaatccc	4080
agcaacacaa	tgaagactaa	gcaaagggga	aagaacaatc	tatcaacaaa	agaaaaatgt	4140
ctctatcctc	tttatcaaca	aactacatat	ttacaaaagt	gaaatgttat	agaacaacta	4200
tgataaacac	agatatatta	tagtaaacc	aagaaagcca	cccacacaga	gctcttaaaa	4260
tcatttgcag	ggttaagcat	agtttaacaa	agtgatgctt	gcaatatatg	cacgtccaca	4320
ctgtctataa	aacaaaaaaa	aagctaaaga	ggattgtaca	gttggtctagt	gatattccatc	4380
ggggatttgt	acctgtaccg	tacccccata	ccaaaatcgg	tgcatcctca	actcccttag	4440
ctcttcagaa	cccatgggta	tgaaaatgag	gtgctcttta	ctcgggtttc	agattctgcc	4500
aatgctgtac	tttccagctg	catgtgcttg	cagatgcaaa	accgcagat	aggagggttg	4560
acatgacaac	tgtacgtcac	agagatgaaa	tgacaaaagta	ttcaacatcc	ttctccaagt	4620
gcttcccaac	aaacaagtga	atcaagaagc	taatgcctat	attaataaaa	atactctgat	4680
ttttacatac	atatgtatca	gcaatgtcta	catattaata	tataacgaat	aggattttta	4740
aagatataac	ctgactgtac	acacagcaat	ctaaaaatcc	cactctacca	cacagggttg	4800
gtatgccatt	tcacttcaac	aaacataaaa	agcttctgcc	ctatagtttt	aagtaaaata	4860
tatccaccaa	atacaacaaa	tttctgcaca	tctgtaaagg	tagatgacat	attatacatg	4920
gacacaacta	acaatcaaat	aaacctctac	atacacacca	aatttttttt	aaaaataaag	4980
gttaacactt	ctattagcaa	aagattgttt	tcaccatatt	ttaacattca	tatatctgtg	5040
atgttcatat	atgtgcatgg	tgggtgtttc	tacataccta	ggaaattcag	atgttattca	5100
agacgcttat	ctattattca	tacagcagtt	aaatgggtctc	cagctctatc	caggcaccca	5160
acagcccaat	aaacacagga	agggacaagc	caccgattag	gttggaataa	gaaaaataaa	5220
tacaaaaatg	tacattttta	tataaaaatt	gttttagacaa	agacaatctc	attttttaga	5280
actccagaac	ccaaaacaaa	accaagaaaa	ccctcaaaat	caaacctcac	aagcacacaa	5340
agatgctaac	tctattggca	acaatgatcc	tactttgcnc	cnccaaaatg	nnctttccca	5400
aaaatatgac	agaaattgta	tattaaaata	tttttgtaat	atttatcttc	tagatacaat	5460
tatatattct	gcatcaacca	aataaaatccc	atttaaacat	ggcatagcaa	ttctactgat	5520
caaactaaag	tgtttatgcc	ct				5542

<210> 129

<211> 2948

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (389)..(412)

<223> a, c, g or t

<400> 129

```

gcctgtgaca ggcacacaggt tagctggctc ccactcgggt ggccgcccga ggatataaat 60
ccggggcgagg gcccctgctg tggctcctct ccctgcacac tcaggagagg gagcttcctt 120
ctaaagacct ttctttttatc tgaagccgca cagcccggca ggctgtgctg acttgggtgga 180
ggcagcagcg gcagagcagc ctgagcagca gcctgagcag gaaacctgct ggggtgggga 240
gggcagggtgt ctgcagcccc tgagaagaag gccctgggtg gcccagacc ctggcatcgt 300
ttcaggggag gtctctagcc gcccagcct gcacacatgt ggcccaagg tgcgcgggt 360
tctggctctcg ctgggagcag gtggcagcnn nnnnnnnnnn nnnnnnnnnn nncggggtgc 420
ccccgcgaag cctggcgctg ccgcccaccc gctattccca cgcgggcatc tgcccaacg 480
acatgaatcc caacctctgg gtggacgcac agagcacctg caggcgggag tgtgagacgg 540
accaggagtg tgagacctat gagaagtgt gcccacacgt atgtgggacc aagagctgcg 600
tgccggcccg ctacatggac gtgaaaggga agaagggccc agtgggcatg cccaaggagg 660
ccacatgtga ccacttcatg tgtctgcagc agggctctga gtgtgacatc tgggatggcc 720
agcccgtgtg taagtgcaa gaccgctgtg agaaggagcc cagctttacc tgcgcctcgg 780
acggcctcac ctactataac cgctgttaca ggctgatttc ccgctgtccc aaaggcatca 840
cactggccgt tgtaacctgc cgctatcact tcactggcc caacaccagc ccccagcac 900
ctgagaccac catgcacccc agcacagcct ccccagagac ccctgagctg gacatggcgg 960
tccctgcgct gctcaacaac cgtgtgcacc agtcgggtcac catgggtgag acagtgagtt 1020
tccctctgtga tgggtggggc cggcccccggc ctgagatcac ctgggagaag cagttggagg 1080
atcgggagaa tgggtgcatg cggcccaacc atgtgcgtgg caacgtgggt gtcaccaaca 1140
ttgcccagct ggtcatcata taacgcccag actgcaggat gctgggatct acacctgcac 1200
gtgcccggaa cgtggctggg tggctgtgaga ggctgatttc ccgctgtcgg atggtcagg 1260
gtcatcaggc atgcagccag catcagagag cagccccaat ggcacggctt tcccgccggc 1320
cgagtgccctg aagccccag acagtgagga ctgtggcgaa gagcagacc gctggcactt 1380
cgatgccag gccacaact gcctgacctt cacttccggc cactgccacc gtaacctcaa 1440
ccactttgag acctatgagg cctgcatgct ggctgcacat agcggggccg tggccgcgtg 1500
cagcctgccc gccctgcagg ggccctgcaa agcctacgag cctcgtggg cttacaacag 1560
ccagacgggc cagtccagc cctttgtcta tgggtggctg gagggcaatg gcaacaactt 1620
tgagagccgt gaggcctgtg aggagtcgtg ccccttcccc agggggaacc agcgtgtcgt 1680
ggcctgcaag cctcggcaga agctcgctac cagcttctgt cgcagcgact ttgtcatcct 1740
gggcccagtc tctgagctga ccgaggagcc tgactcgggc cgcgccctgg tgactgtgga 1800
tgaggtccta aaggtatgaga aaatgggcct caagtccctg ggccaggagc cattggagg 1860
cactctgctt cagtggaact gggcatgccc ctgcccacac gtgaccgtga gcgagatgcc 1920
ggtcatcatc atgggggagg tggacggcgg catggccatg ctgcccggcg atagctttgt 1980
gggcccagtc agtgcccggc gggtcaggaa gcttcgtgag gtcatgcaca agaagacctg 2040
tgacgtcttc aaggagtctt ttggcttgca ctgaagcccc ccacctctc ctgcccctc 2100
cctggccttc ttccacctat ccaccccaat gcctctcagc aaactgggag aggtcagatt 2160
agacaggctt gggacagcag ggaacatca accgacgtgt cacagaaaaa gccacagaag 2220
gtctcagatc agcatctatt ctttgggttc aataaggggt tcatatcttt tttagctgag 2280
ggggacaaga ggagaagtca gtggacacat ggaagtact cgtgaccacc agcttgctca 2340
gatattctcc tctccctcc actggcccca caccctggc tctcccagtc accctccct 2400
agccagtctc ccagcaaggg ttttaagagat ggccgctgtg tgctgggtcac aggaagtgtt 2460
gaatggattg gcttgcaaag ggggtagggt gggagagata ggaggggcca gggactcatg 2520
ggacaccttt cccacagcct cctcgattgc tgtgagcaga ggccactcgg agttaggggc 2580
atgggcaata gcaagctggc ggcagagtc agccagcat atgacttgcc ctgaatggaa 2640
gctgctgaaa cgggtgcctt tgggtgggtg tcggcttgcc tctgaggcca ccacggcacc 2700
agcagaatac gtatttcttc tccttggtg cattgggttg tcgatctagt tcagttcaac 2760
tcagtggatg ttctctgaat gcttaaaact tctggagttt ctgtctgatg gatgggtgtc 2820
tttcatatgc cactggcttc cttggacata gatcagacaa aagccccggg atctgcaatc 2880
tctctgagtc tctgtttcct catctgtctc ctgtctgccc tggatactca ctctcacct 2940
tctgcac 2948

```

<210> 130

<211> 3063
 <212> DNA
 <213> Homo sapiens

<400> 130

```

caggtgtccc accgtgccag akacgctgcc taaactgctt ccagcttctt tttttttttt 60
cccccttct gcaataagtc tgtgatcagc cacgggacag aggcgccagc agcctgcctg 120
tgacaggcat caggttagct ggctccact cgggtggcgc gccaggata taaatccggg 180
cgcgggcccc tgctgtggct cctctccctg cactcagg agaggagct tccttctaaa 240
gacctttctt ttatctgaag ccgcacagcc cggcaggctg tgctgacttg gtggaggcag 300
cagcggcaga gcagcctgag cagcagcctg agcaggaaac ctgctggggt ggggagggca 360
ggtgtctgca gccctgaga agaaggccct ggtgggcccc agaccctggc atcgtttcag 420
gggagggtct tagccgcccc agcctgcacc atgtgggccc caagggtgtcg ccggttctgg 480
tctcgctggg agcaggtggc agcgtgctg ctgctgctgc tactgctcgg ggtgcccccg 540
cgaagcctgg cgctgcgcc catccgctat tcccacgcc gcactctgcc caacgacatg 600
aatcccaacc tctgggtgga cgcacagagc accctgcagg gggagtgtga gacggaccag 660
gagtgtgaga cctatgagaa gtgctgcccc aacgtatgtg ggaccaagag ctgctggtcg 720
gcccgtaca tggacgtgaa aggaagaag ggcccagtgg gcactgcccc ggaggccaca 780
tgtgaccact tcatgtgtct gcagcagggc tctgagtgtg acatctggga tggccagccc 840
gtgtgtaagt gcaaagaccg ctgtgagaag gagcccagct ttacctgcgc ctccgacggc 900
ctcactact ataaccgtg ctacatggat gccaggcct gctccaaagg catcactg 960
gccgttgtaa cctgccgcta tcacttcacc tggcccaaca ccagccccc agcactgag 1020
accaccatgc accccagcac agcctcccca gagaccctg agctggacat ggcggtccct 1080
gcgctgctca acaaccgtgt gcaccagtgc gtcaccatgg gtgagacagt gagtttctc 1140
tgtgatgtgg tgggccggcc ccggcctgag atcacctggg agaagcagtt ggaggatcgg 1200
gagaatgtgg tcatgcggcc caaccatgtg cgtggcaacg tgggtgtcac caacattgcc 1260
cagctgggtca tcatataacg cccagactgc aggatgctgg gatctacacc tgcacgtgcc 1320
cggaacgtgg ctgggtgtcc tgagaggctg atttcccgct gtcggatggt cagggtcat 1380
caggcatgca gccagcatca gagagcagcc ccaatggcac ggctttcccg gcggccagat 1440
gcctgaagcc ccagacagt gaggactgtg gcgaagagca gaccgctgg cacttcgatg 1500
cccaggccaa caactgcctg accttcacct tcggccactg ccaccgtaac ctcaaccact 1560
ttgagacctg tgaggcctgc atgctggcct gcactgagcg gccgctggcc gcgtgcagcc 1620
tgcccgcct gcaggggccc tgcaaagcct acgcgcctcg ctgggcttac aacagccaga 1680
cgggccaagt ccagtcctt gtctatggtg gctgcgagg caatggcaac aactttgaga 1740
gccgtgaggg ctgtgaggag tcgtgcccct tcccagggg gaaccagcg tgtcgggcct 1800
gcaagcctcg gcagaagctc gttaccagct tctgtcgcag cgactttgtc atcctgggcc 1860
gagtctctga gctgaccgag gagcctgact cgggcccgc cctggtgact gtggatgagg 1920
tcctaaagga tgagaaaatg ggctcaagt tcctgggcca ggagccattg gaggtcactc 1980
tgcttcacgt ggactgggca tgcccctgcc ccaacgtgac cgtgagcgag atgccgctca 2040
tcatcatggg ggaggtggac ggcggcatgg ccatgctgcg ccccgatagc tttgtgggcg 2100
catcgagtgc ccgccgggtc aggaagcttc gtgaggatc gcacaagaag acctgtgacg 2160
tcctcaagga gtttcttggc ttgcactgaa gccccccacc cctccctgcc ccctccctgg 2220
ccttcttcca cctatccacc ccaatgcctc tcagcaact gggcgaggtc agattagaca 2280
ggcttgggac agcagggaaa catcaaccga cgtgtcacag aaaaagccac agaaggtctc 2340
agatcagcat ctattctttg ggttcaataa ggggttcata tcttttttag ctgaggggga 2400
caagaggaga agtcagtggg caccatggaag ttactcgtga ccaccagctt gctcagatat 2460
tctcctctc cctcactgg cccacacccc ctggctctcc cagtcaccct cccctagcca 2520
gtctcccagc aagggtttaa gagatggcdg ctgtgtgctg gtcacaggaa gtgttgaatg 2580
gattggcttg caaagggggg aggtggggag agataggagg gccagggac tcatgggaca 2640
cctttccac agcctcctcg attgtgtga gcagaggcca ctggagtta ggggcatggg 2700
caatagcaag ctggcggcag agtccagccc agcatatgac ttgccctgaa tggagctgc 2760
tgaaacgggt gcctttgggt ggtggctggc ttgcctctga ggccaccacg gcaccagcag 2820
aatacgtatt tcttctcctt ggctgcattg gtttgcgat ctagttcagt tcaactcagt 2880
ggatgttctc tgaatgctta aactgtctgg agtttctgtc tgatggatgg tgtgctttca 2940
tatgccactg gcttccctgg acatagatca gcacaaagcc ccgggatctg caatctctc 3000
gagtctctgt ttctcctct gtctcctgtc tgccctggat actcactcct caccttctctg 3060
cac

```

3063

<210> 131
 <211> 904

<212> DNA

<213> Homo sapiens

<400> 131

```

ggagggccag gactcatggg acacctttcc cacagcctcc tcgattgctg tgagcagagg 60
ccactcggag ttaggggcat gggcaatagc aagctggcgg cagagtccag ccagcatat 120
gacttgccct gaattggaagc tgctgaaacg ggtgcctttg ggtggtggtc ggcttgctc 180
tgaggccacc acggcaccag cagaatacgt atttcttctc cttggctgca ctggtttgtc 240
gatctagttc agttcaactc agtggatgtt ctctgaatgc ttactgggtg ccaggaccac 300
agagagatgt tagtcactgc ccagttctta gagccccaac acagataccc tcatcccagg 360
gccccagac acacccctcc gctggactca caactgtctg gagtttctgt ctgatggatg 420
gtgtgctttc atatgccact ggcttccttg gacatagatc agacaaaagc cccgggatct 480
gtttggtagc aggagaaatg aaggaagatg aaaaagcagg cagggaaggg ggtagtaaag 540
gactgagaga ggagggaggt ggctggagaa ggaaaaggaa cattgctcga tgctcccatc 600
tggtggcggc ctgaggaacc cacgggaacc tggaaggagg ctctttgtga gacctgggca 660
aaggatgggg cagctcgtcg atgatttttt tgtgtttcca ggcttcctgt gtgatcctgg 720
ccctccggcc gctagagaga ggattgggaa accccactgt cagctctgca tctgccccca 780
ctaccctcct ctgccctatt ctgtccctgc ccctccaagc tgaagaaggc ccttgtgggg 840
cgctcctcatt tcttcctcaa atataaggag gaagatacca attaaaagct catagtatca 900
atgc 904

```

<210> 132

<211> 442

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (393)

<223> a, c, g or t

<400> 132

```

cactaccata gtggggagggt gtattcataa ctgttgggca tgccaggaaa ttcaggttcc 60
ccaggtagtc tacactggaa atatgggagg agccttggtta ccacctgata gagatgaaag 120
tcccaggtag ctactcaatc tctgtaacac cccagcagga aagttagggt aacttgtag 180
aggctgggtg ggggtggcacc ccaactcagcc tatgctggca taggcagagg tggggacaca 240
gttctttctg tgggtgttag ctggagttag acagttacag tatacaagtt ttctgtctta 300
ctaggttgcc cctttcctgg tctttttgct aaggagagga ggctttattt atttattatt 360
tctatttttg tcttactcac tggcattctg ggntgctggt tcttcagctc caagtctgag 420
atatatggat ccaaaaagaa ac 442

```

<210> 133

<211> 530

<212> DNA

<213> Homo sapiens

<400> 133

```

aatggtcaag aaactttgca tgttaagaaa gtttaagctt tgaaaccttg gaacaacaac 60
tatcatttca catgactctt caccttaaat catctaattg accatgaata ggtgcttttg 120
tcaatattaa atctagaaac atagatatag tatactctga tattaactag gaattataaa 180
gtttataaac tcttgtaaat gtttccattt aaaaatattg tgaaactaaa atgattaata 240
cattaaataa atcaaaattg tatattttta gtctggaagt gcattttcat attccaatta 300
taagtgtgta ttaagcgact gttttcctaa atgtcattat tttatatgaa aaatgccttc 360
attgtctgaa agcattttac tgagttccga ggtttgtgat tggacaaaac tgagcacaat 420
tttctcatct gcaataaatt tactgctaatt ttgttgtaaa gttagctaatt taaataatta 480
ttgtataaaa cgaaatataa tttggtggaa aacgctaaac tggcagatta 530

```

<210> 134

50

<211> 300
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (289)
 <223> a, c, g or t

<400> 134
 gctcgaggct gctaacagag aagccctca tctgtacga ccagtgcaga gaaacgatcc 60
 cctcgaatgc ttcctagtgg agttaagaaa tttttgttg atcgtgcctt tgaactaagg 120
 tcattttaagt atacaacaga tgctcctctg agggaaacag acttataaag tcaggaacac 180
 agaagggacc taatggttta ctagggtgg cgcattaagt tcatagcaat ttaactcctt 240
 tcaatgctaa acaaaacaat gacgcaattt gatgcaaat aaaaacttnt caaaacaatc 300

<210> 135
 <211> 696
 <212> DNA
 <213> Homo sapiens

<400> 135
 cttagaatct ttctctgcag caggctcggt tttctcctca aattcctctg tgtttggcta 60
 agaacaatct gtttttccta cacttgtaa gttgctcgaa attcctaagt cccattcatg 120
 ttctttccaa ggattagcag agcactcctc gcttgtctt catcacactc cctccgcaca 180
 tggggtaaaa attacatttg agtggaaccc tggctatcga tgctgtaaa atggagaact 240
 ttggcgagac tcacttcccc ggggtcaaagt gggaaacagg cctgaaaaac aggcctgagc 300
 atctttaatg atgtgcagaa agagaggggc ctctgcccc acgggcagat gtacacagct 360
 gctaacagag aagccctca tctgtacga ccagtgcaga gaaacgatcc cctcgaatgc 420
 ttcctagtgg agttaagaaa tttttgttg atcgtgcctt tgaactaagg tcattttaagt 480
 atacaacaga tgctcctctg agggaaacag acttataaag tcaggaacac agaagggacc 540
 taatggttta ctagggtgg cgcattaagt tcatagcaat ttaactcctt tcaatgctaa 600
 acaaaacaat gacgcaattt gatgcaaat aaaaacttgt caaaacaatc aaaaaaaaaa 660
 aaaaaaaaaa aaaaaaatc tgcgctcgca agaata 696

<210> 136
 <211> 376
 <212> DNA
 <213> Homo sapiens

<400> 136
 agtctctaaa aatcttgcca taggatttgg tctatacttt taaaaaccac tcttttttca 60
 tgataaagcc cttcaacttg ctctaaaagg caacatagga agagagagac gatgcaggcc 120
 agtcctctcc aaataaggca aaaccagct ttatttttag taatgacttt cccaactgca 180
 agagggcaca agtccatgat ccagcattac agaaaccac caacttcag aaaagtttca 240
 acaactcata aagactcaca tgtgcatgca gacacaaaga cccatttttag ggaagaggcc 300
 ccaagacata gtctgaagcc ccagctgggc acttttctcc atgacaactc ttcagccagc 360
 ctgggacagt gcaacc 376

<210> 137
 <211> 1141
 <212> DNA
 <213> Homo sapiens

<400> 137
 ttggcacgag gagtctctaa aaatcttgcc ataggatttg gtctatactt ttaaaaaacca 60
 ctcttttttc atgataaagc cttcaactt gctctaaaag gcaacatagg aagagagaga 120
 cgatgcaggc cagtctctc caaataaggc aaaaccagc tttattttta gtaatgactt 180

51

```

tcccaactgc aagagggcac aagtccatga tccagcatta cagaaaccca ccaacttcca 240
gaaaagtgtc aacaactcat aaagactcac atgtgcatgc agacacaaag acccatttta 300
gggaagaggc cccaagacat agtctgaagc ccagctggg gccctttctc catgacaact 360
cttcagccag cctggacagt gcaacccttg agtaacccca gctttgctta actgggacaa 420
cccacctctc ctcatcctcc tggagaaatg cagttttgta ttttcctgat gtttgatggg 480
cccgacatca gaggatcctc gaaagtcata ttccctggga aatctgacca aaccgtaaga 540
acgaaaagac tattggctaa ctttgtggag accactgaga gtcagtcct cagcagagga 600
gctggaggga aagagacatt ggaatacttc actgtgattg tccacgcctg cattctcttc 660
atctgtataa actgtggctg gttcacttta accctgagca ggagctgcct atgaaagagg 720
atggctggag tcagatgcct gggcactctt ctggtcaagt cgggagctct cagtgcctgc 780
tgactcatct gtaaaatggg gataacgtca ggatgagcta ataacgcgga agccagaaaag 840
gctgatgcca tctctgtttc caatatgatt tttatggcct ccaagatggg gtccttagaa 900
tctttctctg cagcaggctc gtttttctcc tcaaattcct ctgtgtttgg ctaagaacaa 960
tctgtttttc ctacacttgt caagttgttc gaaattccta atgccattc atgttctttc 1020
caaggattag cagagcactc ctgccttgtc tttcatcaca ctccctccgc acatggggta 1080
aaaattacat ttgagtggaa ccctggctat cgatgcctgt aaaatggaga ctttggcgag 1140
a

```

<210> 138
 <211> 14
 <212> PRT
 <213> Homo sapiens

<400> 138
 Met Gly Tyr Tyr Val Ser Asp Val Leu Leu Asp Leu Val Phe
 1 5 10

<210> 139
 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 139
 Met Phe Leu Ser Ser Val Leu Tyr Cys Ser Leu Leu Ser Tyr Leu His
 1 5 10 15

Leu Ser

<210> 140
 <211> 449
 <212> PRT
 <213> Homo sapiens

<400> 140
 Leu Phe Pro Arg Leu Glu Tyr Gly Gly Thr Ile Leu Ala Tyr Cys Asn
 1 5 10 15

Leu His Leu Pro Gly Ser Ser Asn Pro Pro Thr Ser Ala Ser Gln Val
 20 25 30

Ala Gly Thr Arg Asp Val Cys His His Thr Trp Leu Val Cys Val Cys
 35 40 45

Val Cys Val Cys Val Cys Val Cys Val Cys Val Glu Met Arg Phe His
 50 55 60

Tyr Val Ser Gln Ala Gly Leu Glu Leu Leu Ser Ser Ser Asp Pro Pro

52

65	70	75	80
Ile Ser Ala Ser Gln Ser Ala Gly Ile Ile Gly Ile Ser His Cys Thr	85	90	95
Trp Pro Trp His Asp Ser Phe Ile Ser Pro Gly Ala Glu Leu Pro Thr	100	105	110
Phe Ala Tyr Thr Trp Pro Gly Arg Pro Lys Ile Pro Leu Thr Ile Leu	115	120	125
Leu Leu Tyr Pro Gly Pro Gly Asp Val Leu Val Ala Phe Arg Thr Glu	130	135	140
Leu Tyr Tyr Ala Ser Pro Ser Arg Gln Pro Gly Ala Ser Asp Thr Ala	145	150	155
Arg Glu Ser Trp Gly Asn Gly Ala Val Pro Asp Phe Leu His Lys Glu	165	170	175
Trp Leu Ile Phe Cys Pro Phe Ser Asn Gln Ser His Leu Trp Thr Thr	180	185	190
Lys Ser Lys Trp Ala Glu Val Pro His Pro Gly Arg Arg Ala Glu Leu	195	200	205
Pro Ala Met Lys Glu Gln Lys Ala Ala Asn Glu Asn Ser Gly Ser Val	210	215	220
Thr Glu Pro Ser Ser Ser Ala Ser Ile Leu His Ala Arg Trp Asp Val	225	230	235
Tyr Phe Leu Ile Asn Ala Leu Ile Tyr Phe Leu Arg Gln Ser Leu Arg	245	250	255
Ser Val Ala Gln Ala Gly Val Gln Trp Cys Ser Gly Ala Asp Leu Gly	260	265	270
Ser Leu Gln Pro Leu Pro Pro Gly Phe Lys Ala Phe Pro Cys Leu Ser	275	280	285
Leu Leu Ser Ser Trp Asp Tyr Arg Ser Leu Pro Pro Cys Pro Ala Asn	290	295	300
Phe Phe Val Phe Leu Ile Glu Thr Gly Phe His His Ile Ser Gln Ile	305	310	315
Ser Ile Ser Ala Pro Cys Asp Pro Pro Ala Ser Ala Ser Gln Ser Ala	325	330	335
Gly Ile Thr Gly Met Ser His Cys Ala Gln Pro Asp Val Tyr Tyr Tyr	340	345	350
Val Ser Gly Tyr Ile Gly Lys Gln Asp Arg Cys Tyr Leu Phe Phe Phe	355	360	365
Phe Phe Phe Phe Glu Thr Glu Ser Arg Thr Val Ala Gln Ala Gly Arg	370	375	380
Leu Glu Arg Ser Gly Ala Ile Ser Thr Arg Arg Ser Leu Gln Pro Leu	385	390	395
			400

Pro Pro Gly Leu Lys Arg Phe Ser Cys Leu Ser Leu Leu Ser Ser Trp
 405 410 415

Asp Tyr Arg Cys Thr Pro Pro Arg Leu Ala His Phe Cys Thr Phe Ser
 420 425 430

Arg Asp Gly Val Ser Pro Cys Trp Ser Gly Trp Ser Leu Ser Pro Asp
 435 440 445

Leu

<210> 141
 <211> 11
 <212> PRT
 <213> Homo sapiens

<400> 141
 Met Ile Ala Ile Phe Leu Ser Phe Leu Phe Phe
 1 5 10

<210> 142
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 142
 Met Asp Ala Lys Gln Asn Val Glu Lys Thr Tyr Cys Pro Ala Leu Ser
 1 5 10 15

Gly Ser Phe Gln Asp Ser Met Ile Tyr Trp Glu Arg Ser Asn Ser Leu
 20 25 30

Pro Leu Pro Ala Thr Cys Lys Pro
 35 40

<210> 143
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 143
 Met Asp Gly Phe Val Lys Asp Gln Ala Thr Ser Ser Leu Pro Leu Ala
 1 5 10 15

Thr

<210> 144
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 144
 Met Ala Ser Lys Pro Asn Leu Leu Tyr Ile Leu His Tyr Cys Val Pro
 1 5 10 15

Asp Thr Ala Asn Ser Ile Asn Glu
20

<210> 145
<211> 20
<212> PRT
<213> Homo sapiens

<400> 145
Met Ser Cys Ser Ser Ser Thr Gly Ala Gly Lys Tyr Asn Leu Lys Gly
1 5 10 15

Glu Ala Asn Leu
20

<210> 146
<211> 107
<212> PRT
<213> Homo sapiens

<400> 146
Tyr Tyr Phe Tyr Tyr Tyr Phe Phe Leu Arg Glu Ser Leu Thr Leu Ser
1 5 10 15

Leu Gly Leu Glu Cys Ser Gly Val Thr Met Ala His Gln Thr Ile Asn
20 25 30

Ile Pro Gly Ser Ser Asn Ser Pro Val Val Val Gly Thr Thr Gly Ala
35 40 45

Cys His Asn Ala Trp Leu Ile Phe Val Phe Leu Val Glu Thr Gly Leu
50 55 60

His His Val Gly Gln Ala Gly Leu Gly Leu Leu Ala Ser Ser Asp Leu
65 70 75 80

Ser Ala Leu Ala Ser Pro Ser Ala Gly Ile Ile Gly Leu Ser His Cys
85 90 95

Thr Gln Gln Lys Thr Asn Phe Leu Lys Gln Asn
100 105

<210> 147
<211> 18
<212> PRT
<213> Homo sapiens

<400> 147
Met Arg Ser Asn Phe Lys Lys Asn Ile Pro Ser Leu Glu Leu Phe Asn
1 5 10 15

Met Ser

<210> 148
<211> 99

55

<212> PRT

<213> Homo sapiens

<400> 148

Leu Phe Ser Phe Ala Arg Gln Asp Val Ser Met Leu Pro Arg Leu Glu
 1 5 10 15
 Tyr Ser Gly Gly Ile Ile Ala His Cys Lys Leu Asp Val Leu Asp Ser
 20 25 30
 Ser Glu Leu Thr Ala Leu Thr Ser Gln Ile Ala Gly Thr Thr Gly Val
 35 40 45
 His His His Ala Arg Leu Ile Phe Thr Met Phe Met Gln Met Gly Ser
 50 55 60
 Cys Ser Val Ala Gln Ala Cys Leu Lys Leu Leu Ala Ser Asp Asp Pro
 65 70 75 80
 Pro Ala Phe Gly Ser Gln Ser Ala Gly Ile Ala Asp Val Ala His His
 85 90 95
 Ala Gln Pro

<210> 149

<211> 64

<212> PRT

<213> Homo sapiens

<400> 149

Met Ser Val Ser Val Leu Pro Val Gln Pro Pro Thr Gly Leu Leu Trp
 1 5 10 15
 Gly Arg Ser Pro Pro Gly Ser Pro Ala Glu Leu His Gly Leu Pro Cys
 20 25 30
 Leu Thr Arg Asp Asn Arg Asp Phe Gly Ser Pro Ser Ala Asp Ala Phe
 35 40 45
 Val Leu Phe Leu Ile Arg Ser Arg Thr Arg Val Gly Arg Arg Val Met
 50 55 60

<210> 150

<211> 23

<212> PRT

<213> Homo sapiens

<400> 150

Met Val Glu Ser Gly Ile Glu Pro Glu Asn Ser Asp Ser Arg Leu Ser
 1 5 10 15
 Cys Phe Ser His Arg Ala Val
 20

56

<210> 151
<211> 27
<212> PRT
<213> Homo sapiens

<400> 151
Met Ile Gln Arg Leu Leu Arg Gly His Asn Cys Ile Ser Ile Pro Asn
1 5 10 15
Leu Phe Tyr Asn Glu Arg Ile Tyr Arg Ile His
20 25

<210> 152
<211> 26
<212> PRT
<213> Homo sapiens

<400> 152
Met Pro Ser Ala Trp Lys Val Glu Asp Ser Gly Ile Arg Glu Arg Phe
1 5 10 15
Arg Pro Gly Glu Met Glu Gly Ser Gly Thr
20 25

<210> 153
<211> 16
<212> PRT
<213> Homo sapiens

<400> 153
Met Gln Val Trp Ser Gly Ile Phe Pro Asp Arg Gly Cys Cys Ser Cys
1 5 10 15

<210> 154
<211> 61
<212> PRT
<213> Homo sapiens

<400> 154
Met Phe Met Trp His Arg Val Ala Asn Cys Leu Ser Leu Phe Val Ser
1 5 10 15
Gln Asn Asp Phe Ala Asp Val Leu Gly Gln Ala Ser Pro Gly Trp Gln
20 25 30
Pro Gly Ala Ala Val Lys Phe Ser Leu Thr Asn Ser Leu Pro Pro Phe
35 40 45
Pro His His Gly Thr Leu Val Leu Cys Val Thr Thr Val
50 55 60

<210> 155
<211> 69
<212> PRT
<213> Homo sapiens

<400> 155

57

Met Pro Cys Trp Lys Leu Leu Met Asn Arg Ala Trp Ser Leu Thr Leu
 1 5 10 15
 Gly Gly Gln Val Ile Tyr Arg Gly Asn Asp Asn Val Asn Pro Gly Pro
 20 25 30
 Trp Gly Ala Gly Ser Val Val Lys Glu Thr Gln His Thr Gln Gly Trp
 35 40 45
 Asp Pro Thr Gln Ala Lys Glu Gly Ser Thr Pro Ser Pro Asp Val Cys
 50 55 60
 Trp Asn Lys Glu Lys
 65

<210> 156

<211> 51

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (7)

<400> 156

Met Lys Lys Lys Arg Phe Xaa Tyr Asn Ile Lys Ile Leu Val Asn Ser
 1 5 10 15
 Trp Leu Glu Leu Tyr Ser Glu Ile Thr Val Phe Lys Lys Asp Arg Pro
 20 25 30
 Leu Pro Leu Ser Leu Trp Leu Met Ala Leu Ile Ile Thr Arg Ile Pro
 35 40 45
 Lys Met Ser
 50

<210> 157

<211> 126

<212> PRT

<213> Homo sapiens

<400> 157

Met Lys Leu Leu Ser Arg Lys Met Trp His Ser Leu Leu Gly Gly Gly
 1 5 10 15
 Trp Gly Gly Gly Lys Arg Glu Gly Arg Cys Pro Gln Leu Pro Pro Arg
 20 25 30
 Ser Ile Asn Lys Lys Arg Ile Asp Pro Pro Ala Pro Phe Asn Ser Pro
 35 40 45
 Pro Glu Leu Pro Pro Asn Ser Val Lys Thr Cys Gly Phe Asp Tyr Ser
 50 55 60
 Asp Glu Asn Asn Gly Cys Ser Val Glu Ile Cys Arg Ala His Thr His
 65 70 75 80
 Met Ile Ser Lys Ser Asn Ser Val Ala Thr Val Pro Ile Arg Lys Thr

58

85

90

95

His Gln Ala His Lys Arg Asp Pro Phe Ile Gln Arg Ser Leu Cys Ile
 100 105 110

Pro Ile Ser Thr His Ser Thr Cys Ile Phe Lys Pro Ile Ser
 115 120 125

<210> 158

<211> 84

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (21)

<220>

<221> UNSURE

<222> (35)

<220>

<221> UNSURE

<222> (45)

<220>

<221> UNSURE

<222> (48)

<220>

<221> UNSURE

<222> (52)

<220>

<221> UNSURE

<222> (58)

<220>

<221> UNSURE

<222> (61)

<400> 158

Met Lys Arg Pro Pro Val Leu Leu Gln Glu Lys Pro Pro Glu Gly Asn
 1 5 10 15

Gly Ala Val Ala Xaa Trp Pro Val Val Thr Pro Arg Arg Gly Arg Gly
 20 25 30

Gln Gly Xaa Leu Gly Pro Gln Asn Ile Val Pro Val Xaa Ser Phe Xaa
 35 40 45

Ala Gly Leu Xaa Leu Leu Arg Ser Leu Xaa Gly Ser Xaa Leu Asn Ser
 50 55 60

Leu Leu Ser Ala Ser Trp Ala Val Val Ser Gly His Arg Leu Leu Leu
 65 70 75 80

Thr Ser Pro Pro

<210> 159
 <211> 23
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (20)

<400> 159
 Met Asp Ser Ala Lys Leu Gly His Ile Cys Tyr Thr Asp Asp Thr Ser
 1 5 10 15
 Leu Asp Val Xaa Ala Gln Thr
 20

<210> 160
 <211> 50
 <212> PRT
 <213> Homo sapiens

<400> 160
 Met Ile Asn Phe Ala Phe Val Val Cys His Lys Thr Thr Val Thr Val
 1 5 10 15
 Ser Leu Gln Leu Lys Ile Ile Gly Tyr Ala Thr Pro Glu Gly Asn Gln
 20 25 30
 His Ser Lys Cys Ile Pro Ser Ile Val Phe Ile Ile Cys Glu Arg Met
 35 40 45
 Ser His
 50

<210> 161
 <211> 57
 <212> PRT
 <213> Homo sapiens

<400> 161
 Met Met Pro Thr Asp Asn Leu Leu Met Ile Ser Ser Ile Leu Lys Asp
 1 5 10 15
 Val Cys Lys Thr Gln Pro Leu Arg Lys Asp Ser Tyr His Cys Ser His
 20 25 30
 Arg His Pro Pro Gln Ser Tyr Thr Phe Pro Phe His Pro Pro Lys Gln
 35 40 45
 Ile Ile Gln His Ile Tyr Phe Ile Leu
 50 55

<210> 162
 <211> 10
 <212> PRT
 <213> Homo sapiens

61

20 25 30
 Ser Ser Leu Val Ser Ser Leu Asp Leu Asn Glu Asn Ile Ser Val Tyr
 35 40 45
 Phe Thr Thr Lys Tyr Glu Leu Ala Ser Gly Cys Ala Leu Phe Tyr Phe
 50 55 60
 Tyr Thr Glu Cys Phe Lys Thr Asn
 65 70

<210> 166
 <211> 57
 <212> PRT
 <213> Homo sapiens

<400> 166
 Met Ser Cys Ser Val Leu Leu Arg Lys Cys Tyr Asn Arg Ala Asp Gln
 1 5 10 15
 Phe His His Val Phe Ile Ile Thr Ile Leu Arg Trp Ala Leu Asn Thr
 20 25 30
 Ala Gln Gln Ala Cys His Phe His Leu Ile Ser Ser Ala Thr His Phe
 35 40 45
 Leu Leu Glu Leu Ala Ser Ser Asn Leu
 50 55

<210> 167
 <211> 121
 <212> PRT
 <213> Homo sapiens

<400> 167
 Met Thr Pro Leu Leu Pro Gly Gly Glu Gln Leu Arg Glu Asn Trp Arg
 1 5 10 15
 Ala Gln Thr Thr Gln Leu Gly Arg Gly Gly Gly Leu Met Glu Pro Arg
 20 25 30
 Ala Leu Arg Ala Ser Pro Gly Ser Ser Pro Pro Ala Pro Pro Leu Pro
 35 40 45
 Glu Ser Pro Ser Leu Ser Trp Cys Ala Gly Arg Thr Cys Ala Ala Ala
 50 55 60
 Ala Gly Gly Gly Cys Thr Ser Gly Arg Glu Leu His Ala His Trp Glu
 65 70 75 80
 Gln Pro Met His Arg Pro Pro Arg Cys Ala Gln Val Ser Gly Ala Ser
 85 90 95
 Gly Lys Glu Glu Lys Ala Ala Val Ser Ala Leu Ser Leu Ser Leu Met
 100 105 110
 Pro Val Trp Asn Pro Thr Asp Glu Leu
 115 120

<210> 168
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 168
 Met Gly Glu Val Val Tyr Leu Phe Lys Val Pro Cys Leu Val Tyr Thr
 1 5 10 15

His

<210> 169
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 169
 Met Ser Asn Tyr Tyr Ser Phe Ile Ile Asn Leu Asn Ser Phe Gln Ile
 1 5 10 15

Arg Ala Thr Pro Ser Pro Cys Pro Leu Phe Gln Glu Tyr Phe Gly Ser
 20 25 30

Ser Trp Phe Phe Val Ser Pro Tyr Asp Asp Phe Thr Ile His Leu
 35 40 45

<210> 170
 <211> 33
 <212> PRT
 <213> Homo sapiens

<400> 170
 Met Lys Ala Ile Gln Ile Glu Glu Phe Phe Ala Ser Leu Leu Thr Gly
 1 5 10 15

Pro Gly Val Leu Asp Asn Phe Leu Ser Lys Glu Glu Lys Asn Ile Phe
 20 25 30

His

<210> 171
 <211> 49
 <212> PRT
 <213> Homo sapiens

<400> 171
 Met Asp Ala Cys Leu Gly Asp Cys Gln Pro Gln Gly Arg Ser Ile Asp
 1 5 10 15

Leu Lys Tyr Glu Gln Thr Asp Asp Phe Ile Ile Met Thr Leu Ala Gln
 20 25 30

Asn Arg Asn Phe Gly Thr Glu Lys Asn Lys His Met Glu Phe Leu Lys
 35 40 45

63

Gly

<210> 172

<211> 56

<212> PRT

<213> Homo sapiens

<400> 172

Met	Ser	Leu	Lys	His	Asn	Asn	Ile	Ile	Phe	Tyr	Ser	Gln	Glu	Glu	Leu
1				5					10					15	

Ile	His	Asp	Arg	Ile	Ile	Ser	Leu	Ala	Ile	Leu	Tyr	Ser	Tyr	Phe	Val
			20					25					30		

Leu	Phe	Ser	Ser	Phe	Pro	Leu	Pro	Phe	Asp	Asp	Gln	Phe	Leu	Tyr	Lys
		35					40					45			

Thr	His	Arg	Tyr	Ile	Pro	Phe	Ile
	50					55	

<210> 173

<211> 79

<212> PRT

<213> Homo sapiens

<400> 173

Met	Gly	Glu	Ile	Gln	Val	Asp	Leu	Asn	Cys	His	His	Gln	Ser	Arg	Pro
1				5					10					15	

Arg	Arg	Arg	Leu	Leu	Ser	Arg	Met	Tyr	Thr	Trp	Pro	Leu	Phe	Ala	Val
			20					25					30		

Ala	Val	Leu	Leu	Leu	Arg	Gly	Glu	Pro	Ile	Tyr	Val	Cys	Leu	Phe
		35				40					45			

Leu	Leu	Ser	Leu	Ala	Ala	Gln	Gln	Asn	Pro	Val	Ile	Tyr	Met	Asn	Lys
	50					55					60				

Phe	Leu	Glu	Val	Lys	Arg	Asp	Glu	Lys	Phe	Thr	Lys	Ser	Pro	Thr
65					70					75				

<210> 174

<211> 30

<212> PRT

<213> Homo sapiens

<400> 174

Met	Val	Leu	Lys	Gly	Met	Asn	Ile	Thr	Glu	Ile	Glu	Cys	Phe	Leu	Gln
1				5					10					15	

Val	Glu	Arg	Leu	His	Ser	Leu	Ala	Gly	Thr	Phe	Cys	Pro	Ile
			20					25				30	

<210> 175

<211> 73

<212> PRT

64

<213> Homo sapiens

<400> 175

Met Ala Gly Ala Gly Gly Gln His His Pro Pro Gly Ala Ala Gly Gly
1 5 10 15

Ala Ala Ala Gly Ala Gly Ala Ala Val Thr Ser Ala Ala Ala Ser Ala
20 25 30

Gly Pro Gly Glu Asp Ser Ser Asp Ser Glu Ala Glu Gln Glu Gly Pro
35 40 45

Gln Lys Leu Ile Arg Lys Val Ser Thr Ser Gly Gln Ile Arg Thr Lys
50 55 60

Gly Phe Ile Met Leu Ala Arg Leu Val
65 70

<210> 176

<211> 33

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (22)

<400> 176

Met Glu Ile Trp Leu Leu Ala Leu Ala Phe Lys Lys Leu Ser Arg Arg
1 5 10 15

Phe Tyr Val Gln Pro Xaa Leu Gly Thr Thr Val Leu Gly Asn Ile Arg
20 25 30

Arg

<210> 177

<211> 22

<212> PRT

<213> Homo sapiens

<400> 177

Met Leu Phe Ser Ile Leu Pro His Lys Gly Tyr Ile Leu Lys Asp Ile
1 5 10 15

Trp Leu Leu Asn Leu Asn
20

<210> 178

<211> 45

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (21)

65

<400> 178

Met Leu Leu Lys Gly Ser Asn Ser Lys Val Ser Arg Glu Tyr Ser Ala
 1 5 10 15

Thr Phe His Lys Xaa Thr Glu Gln Ser Ser Arg Asn Phe Phe Arg Ala
 20 25 30

Gly Ile Ala Leu Pro Pro Arg Ile Leu Thr Arg Phe Ser
 35 40 45

<210> 179

<211> 38

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (21)..(22)

<400> 179

Met Val Ala Thr Leu Trp Leu Asn Asn Phe Phe Arg Asn His Lys Asn
 1 5 10 15

Ala Val Lys Asp Xaa Xaa Lys Arg Leu Lys Ala Ile Leu His Ser Leu
 20 25 30

Val Tyr Met Lys Gly Asn
 35

<210> 180

<211> 65

<212> PRT

<213> Homo sapiens

<400> 180

Ser Trp Cys Ser Gly Leu Met Pro Ser Val Leu Asn Ser Ile Ser Cys
 1 5 10 15

Val Pro Gly Lys Gly Arg Gly His Ser Leu Glu Trp Phe Pro Gly Glu
 20 25 30

Lys Ser Gln Ser Asn Leu Cys Ser Ser Phe Leu Asn Lys Asn Arg Arg
 35 40 45

Gln Asn Lys Gly His Arg Asp Lys Gly Leu Leu Thr Arg Leu Ala Asn
 50 55 60

Gln
 65

<210> 181

<211> 12

<212> PRT

<213> Homo sapiens

<400> 181

Met Ala Phe Gly Ile Tyr Gln Cys Leu Gly Met Phe
 1 5 10

<210> 182
 <211> 23
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (21)

<400> 182
 Met Leu Leu Thr Pro Gln Pro Trp Phe Phe Lys Val Ile Phe Val Asn
 1 5 10 15
 Tyr Lys Val Arg Xaa Tyr Lys
 20

<210> 183
 <211> 29
 <212> PRT
 <213> Homo sapiens

<400> 183
 Met Tyr Lys Ile Arg Lys Ser Arg Pro Glu Glu Asp Ser His Cys Leu
 1 5 10 15
 Gln Arg Thr Ala Lys Gly Lys Gly Phe Lys Ile Phe Asn
 20 25

<210> 184
 <211> 58
 <212> PRT
 <213> Homo sapiens

<400> 184
 Met Leu Phe Leu Val Ser Ala Ala Leu Ser Ser Ser Leu Thr Asp Asn
 1 5 10 15
 Cys Arg Ala Gln Val Gly Arg Lys Asn Ser Val Cys Leu Leu Gly Ser
 20 25 30
 Ala Ser Ala Pro Val Ser Asn Thr Gly Val Thr Gly Gly Leu Leu Asn
 35 40 45
 Val Lys Tyr Lys Gly Ser Ser Phe Ser Leu
 50 55

<210> 185
 <211> 21
 <212> PRT
 <213> Homo sapiens

<400> 185
 Met Gln Cys Gln Gln Leu Gly Phe Ser Glu Ile Ile Ser Arg Leu Gln
 1 5 10 15
 Ser Asn Gln Ile Ser

20

<210> 186
 <211> 16
 <212> PRT
 <213> Homo sapiens

<400> 186
 Met Lys Val Glu Arg Gln Phe Glu Ala Arg Ser Leu Thr Asp Ser Leu
 1 5 10 15

<210> 187
 <211> 104
 <212> PRT
 <213> Homo sapiens

<400> 187
 Gln Ile Val Asn Phe Phe Phe Phe Leu Arg Trp Ser Leu Ala Leu Val
 1 5 10 15
 Thr Gln Ala Gly Val Gln Trp Pro Asp Leu Ser Ser Leu Gln Pro Leu
 20 25 30
 Pro Pro Gly Phe Lys His Phe Ser Cys Leu Ser Leu Pro Ser Ser Ala
 35 40 45
 Asp Leu Ser His Val Pro Leu Cys Pro Ala Asn Phe Ala Asn Phe Phe
 50 55 60
 Val Glu Met Gly Ser His Cys Val Thr Gln Ala Gly Leu Ala Val Leu
 65 70 75 80
 Ala Ala Ser Asp Ser Leu Thr Leu Ala Pro Gln Ser Ala Gly Ile Ile
 85 90 95
 Gly Met Ser His Gly Ala Cys Pro
 100

<210> 188
 <211> 41
 <212> PRT
 <213> Homo sapiens

<400> 188
 Met Asp Arg Asp Leu Arg Pro Ala Pro Arg Asp Thr Lys Asp Gly Ser
 1 5 10 15
 Ser Val Ala Ser Ser Pro Asn Ser Ile Cys Pro Cys Leu Ala Arg Cys
 20 25 30
 Arg Glu Asp Phe Pro Thr Gln Glu Lys
 35 40

<210> 189
 <211> 39
 <212> PRT
 <213> Homo sapiens

68

<400> 189

Met Cys Leu Lys Gln Ile Leu Leu Glu Phe Pro Lys Arg Leu Asp Ile
 1 5 10 15

Ile Asn Thr Phe Met Tyr Thr Trp His Pro Thr Arg Ala Val Cys Phe
 20 25 30

Tyr Lys Lys Trp His Lys Asn
 35

<210> 190

<211> 53

<212> PRT

<213> Homo sapiens

<400> 190

Phe Ser Ser Leu Met Lys Val Ile Thr Asp Trp Ala Gln Trp Leu Thr
 1 5 10 15

Pro Val Ile Pro Val Leu Trp Glu Val Ala Val Val Gly Ala Leu Glu
 20 25 30

Ala Arg Ser Leu Arg Pro Ala Trp Glu Thr Ala Thr Pro Phe Pro Phe
 35 40 45

Ala Lys Lys Lys Lys
 50

<210> 191

<211> 44

<212> PRT

<213> Homo sapiens

<400> 191

Met Lys Ala Leu Cys Arg Leu Ser Val Leu Gln Met Leu Val Met Gly
 1 5 10 15

Met Val Val Met Arg Lys Val Met Pro Val Thr Met Arg Arg Gly Asp
 20 25 30

Ala Val Asn Ser Ile His Pro Val Leu Gly Lys Tyr
 35 40

<210> 192

<211> 53

<212> PRT

<213> Homo sapiens

<400> 192

Met Ser Leu Ser Leu Asp Ser Leu Ser Ser Ile Cys Leu Ile Val Asp
 1 5 10 15

Leu Leu Asn Phe Ser Tyr Met Glu Phe Thr Glu Arg Leu Glu Cys Glu
 20 25 30

Asp Gln His Phe Ser Ser Asn Leu Val Ser Phe Gln Ala Met Ile Ser
 35 40 45

Ser Asp Ile Leu Pro
50

<210> 193
<211> 124
<212> PRT
<213> Homo sapiens

<400> 193
Met Arg Phe Leu Leu Pro Ala Ala Glu Lys Arg Lys Glu Asn Ser Ala
1 5 10 15
Gly Ala Pro Leu Ala Ser Pro Arg Val Thr Thr Met Phe Ser His Asp
20 25 30
Arg Gln Thr Gly Ala Leu Leu Leu Cys Asp Pro Pro Arg Ala Ala Glu
35 40 45
Ser Ile Leu Ile His Leu Gly Thr Pro Ala Gln Glu Glu Pro Gly Pro
50 55 60
Ser Pro Phe Arg Asp Val Asp Pro Leu Arg Gly Glu Phe Ser Ser Val
65 70 75 80
Asp Ser Asp Leu Leu Arg Leu Thr Ser Leu Gly Asn Pro Ala Ile Ala
85 90 95
Val Gly Asn Gln Val Ala Ala Trp Ala His Met Ala Ser Arg Arg Leu
100 105 110
Arg Leu Thr Ser Lys Arg His Ser Gln Arg Arg Lys
115 120

<210> 194
<211> 44
<212> PRT
<213> Homo sapiens

<400> 194
Met Phe Gln Arg Ile Ser Val Phe Ser Pro Ala Ile Thr Asn Lys Ser
1 5 10 15
Ser Gly Phe Ala Val Pro Pro Cys Lys Asn Tyr Lys Met Ala Glu Asn
20 25 30
Asn Ala Cys Phe Ile Ile Leu Val Lys Trp Ser Thr
35 40

<210> 195
<211> 27
<212> PRT
<213> Homo sapiens

<400> 195
Met Val Arg Arg His Ile Gly Ser Ala Val Arg Trp Pro Leu Phe Phe
1 5 10 15

70

Ser Asn Trp Ser Pro Tyr Ala Ser Cys Cys Asn
 20 25

<210> 196
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 196
 Met Thr Lys Ile Cys Phe Leu Asn Pro Thr Leu Ala Phe Lys Lys Ile
 1 5 10 15

Gln Ser Lys Ile Phe Arg Leu Phe Leu Lys Asp Glu Lys Ala Ala
 20 25 30

<210> 197
 <211> 25
 <212> PRT
 <213> Homo sapiens

<400> 197
 Met Tyr Met His Tyr Arg Asp Arg Lys Thr Gln Phe Asn Ile Lys Asn
 1 5 10 15

Asn Ile Ser Leu Leu Asn Asn Ala Val
 20 25

<210> 198
 <211> 82
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (80)

<400> 198
 Met Gly Met Val Ala Gly Ala Pro Thr Ala Trp Asn Pro Glu Asp Lys
 1 5 10 15

Gly Cys Ile Leu Leu Gly Arg Gln Ser Tyr Glu Leu Asp Ala Met Trp
 20 25 30

Pro Leu Gly Ala Leu Cys Arg Thr Ala Thr Ile Pro Ala Leu Leu Asp
 35 40 45

Gly Glu Ser Glu Ala Leu Arg Ser Asp Glu Asn Gln Trp Gln Ser Gln
 50 55 60

Met Tyr His Phe Ser His Thr Leu Thr Phe Phe Cys Phe Val Pro Xaa
 65 70 75 80

Phe Phe

<210> 199
 <211> 46

71

<212> PRT

<213> Homo sapiens

<400> 199

Met Pro Leu Arg Ser Lys Leu Val Asn Ile His Leu Phe Leu Thr Thr
 1 5 10 15

Ala Thr Val Phe Ser Leu Tyr Thr Asn Tyr Thr Ala Ser Lys Phe Ser
 20 25 30

Ser Phe Pro Ala Ser Asn Gln Glu Phe Asn Met Glu Val Gln
 35 40 45

<210> 200

<211> 74

<212> PRT

<213> Homo sapiens

<400> 200

Met Gln Val Gln Arg Pro Thr Ser Trp Gly His Ile Ser Thr Ala Phe
 1 5 10 15

Arg Ala Ala Pro Glu Ser Ser Arg Ser Phe Leu Ser Leu Leu Gln Thr
 20 25 30

Phe Phe Glu Lys Trp Thr Phe His Pro His Val Pro Ser Val Trp Leu
 35 40 45

Arg Lys Ser Thr Ser Gly Pro Trp Glu Gly Pro Gly Lys Pro Phe Pro
 50 55 60

Leu Ser Leu Trp Cys Val Gly Ile Asn Leu
 65 70

<210> 201

<211> 150

<212> PRT

<213> Homo sapiens

<400> 201

Met Asn Gly Lys Thr Gln Cys Lys Ala Pro Asn Asp Ser Val Arg Ser
 1 5 10 15

Val Val Gly Arg Thr Asn Thr Trp Ile His Arg Thr Glu Ile Asp Asn
 20 25 30

Leu Ala Cys Asp Glu Leu Lys Ala Asp Ile Leu Asn Trp Trp Arg Lys
 35 40 45

Glu Tyr Leu Leu Ile Ile Gly Ile Thr Ala Phe Leu Phe Leu Phe Arg
 50 55 60

Gly Ala Ile Leu Lys Asp Lys Gln Pro Thr Gly Lys Leu Gly Gln His
 65 70 75 80

Asn Thr Asn Arg Gln Cys Thr Val Glu Ile Tyr Lys Trp Pro Ile Asn
 85 90 95

Met Glu Met Phe Asp Phe Val Arg Asn Gln Gly Asn Ser Ser Glu Asn

72

100 105 110
 Lys Val Leu Ser Ile Thr Arg Leu Val Lys Thr Lys Gln Asn Asn Leu
 115 120 125
 Ser Ile Leu Ile Pro Leu Thr Val Gly Lys Gly Leu Glu Lys Trp Val
 130 135 140
 Leu Leu Trp Arg Val Asn
 145 150

<210> 202
 <211> 33
 <212> PRT
 <213> Homo sapiens

<400> 202
 Met Ala Ala Arg Leu Pro Thr Leu Thr Arg Tyr Lys Phe Ser Ser Leu
 1 5 10 15
 Gly Ser Trp Tyr Lys Ser Gln Pro Phe Gln Leu Val Met Asn Glu Arg
 20 25 30

Ala

<210> 203
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (9)

<220>
 <221> UNSURE
 <222> (23)

<220>
 <221> UNSURE
 <222> (42)

<400> 203
 Met Gln His His Phe Ser Leu His Xaa Pro Cys Arg Asp Leu Pro Gly
 1 5 10 15

Ala Gln Lys Lys Lys Asp Xaa Ile Cys Cys Ser Gln Glu Met Leu His
 20 25 30

Ile Val His Leu Pro Ala Ser Tyr Arg Xaa Tyr Lys Tyr Glu Ser Thr
 35 40 45

Asn Ser Leu Gly Phe Asn Asn Val Thr Tyr Ile Tyr His Lys Val Ala
 50 55 60

Ile Pro Asp His
 65

<210> 204
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 204
 Met Thr Ala Ser Leu Cys Leu Gln Pro Lys Pro Leu Leu Ser Thr Asn
 1 5 10 15
 Pro Tyr Ala His Gly Ala Glu Thr Ala Gln Pro Ser Val Lys Glu Pro
 20 25 30
 Gly Phe

<210> 205
 <211> 115
 <212> PRT
 <213> Homo sapiens

<400> 205
 Leu Ala Ala Ile Tyr Gly Phe Leu Ser Phe Phe Phe Phe Phe Phe
 1 5 10 15
 Ala Asp Lys Val Ser Leu Ser Pro Arg Leu Glu Ala Cys Asn Gly Thr
 20 25 30
 Ile Thr Ala His Gly Ser Phe Asp Phe Leu Gly Ser Gly Asp Pro Pro
 35 40 45
 Thr Ser Ala Ser Ala Ile Ala Gly Thr Gly Ala His His His Ile Ala
 50 55 60
 Leu Leu Phe Val Phe Phe Val Glu Val Gly Ser Arg Tyr Val Ala Gln
 65 70 75 80
 Ala Ala Leu Gln Leu Leu Arg Ser Gly Asp Leu Pro Ala Ser Ala Ser
 85 90 95
 Gln Ser Thr Gly Ile Thr Gly Thr Ser His Cys Ser Trp Pro Tyr Met
 100 105 110
 Val Leu Phe
 115

<210> 206
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 206
 Met Phe Ala Ser Tyr Lys Leu Asn Asn Tyr Ser Tyr Pro Val Leu Val
 1 5 10 15
 Leu Tyr Ala Thr Leu Phe Pro His His Met Ile Phe
 20 25

74

<210> 207
<211> 68
<212> PRT
<213> Homo sapiens

<400> 207
Met Ser Leu Ser Pro Ile Tyr Phe Asn Ala Ser Phe Val Ile Ser Glu
1 5 10 15
Tyr Met Ser Asn Phe Tyr Phe Asn Ser Thr Cys His Leu Cys Tyr Glu
20 25 30
Asp Trp Lys Pro Ser Phe Ser Pro Gly Leu Gly Glu Ala Lys Cys Phe
35 40 45
Thr Tyr Leu Glu Cys Leu Cys His Ser Asn Phe Gln Leu Val Cys Asn
50 55 60
Cys Ser Phe Asn
65

<210> 208
<211> 39
<212> PRT
<213> Homo sapiens

<400> 208
Met Asn Glu Tyr Val Asn Glu Cys Leu Asn Glu Trp Ser Gly Met Asn
1 5 10 15
Pro Val Ser Pro Val Leu Cys Pro Pro Leu Ile His Ser Val Thr Leu
20 25 30
Gly Arg Thr Phe Asn His Ser
35

<210> 209
<211> 45
<212> PRT
<213> Homo sapiens

<400> 209
Met Pro Phe Pro Ser His Ser Leu Leu Leu His Phe Phe Pro Pro Glu
1 5 10 15
Arg Leu Ser Ser Gly Pro Tyr Glu Ile Ala Ser Ile Gln Leu Phe Phe
20 25 30
Ile Leu Lys Gly Asp Asn Ser Ile Ser Phe Asn Leu Asn
35 40 45

<210> 210
<211> 70
<212> PRT
<213> Homo sapiens

<400> 210
Leu Gly Ser Leu Gln Pro Pro Pro Gly Phe Lys Ala Phe Ser Cys

1 5 10 15

Lys Val Leu Gly Leu Gln
65 . 70

Asn Arg Ile Ala Asn Val Ser Ser
20

Lys Thr Lys Glu Gly Ile Thr Pro Asp Phe Glu Ile His Tyr Lys Thr
130 135 140

76

Val Val Thr Lys Thr Val Cys His Leu Asn Lys Asn Arg Asp Ile Gly
 145 150 155 160
 Gln Trp Ser Arg Arg Lys Arg Glu Gln Lys Tyr Ile Ser Val Phe Thr
 165 170 175
 Ala Asn Ala Phe Ala Ile Gln Val Thr Phe Phe Phe Ala Gly Lys Asn
 180 185 190
 Ser Ile Phe Asn Lys Ala Cys Leu Glu Asn Phe Met Ser Thr Cys Arg
 195 200 205
 Lys Lys Lys Ala Asp Pro His Leu Thr Pro Tyr Val Lys Ile Asn Ser
 210 215 220
 Lys Ala Ile Ser His Leu Asn Val Arg Pro Lys Thr Leu Lys Leu Leu
 225 230 235 240
 Tyr Gln Lys Ile Glu Ala Lys Pro His Asn Ile Gly Leu Gly Ser Lys
 245 250 255
 Phe Phe Asp Leu Thr Ala Ile Ser Gln Asp Thr Lys Gly Arg Thr Ser
 260 265 270
 Gln Ser Asp His Phe Lys Leu Lys Ser Cys Cys Thr Glu Ser Asp Thr
 275 280 285
 Ala Thr Glu Val Thr Thr Lys Lys Arg Glu Lys Ile Phe Ala Asn Tyr
 290 295 300
 Thr Cys Asp Lys Gly Leu Ile Ala Lys Ile Tyr Thr Lys Leu Lys Ala
 305 310 315 320
 Gln Tyr Asn Lys Asn Lys Ala Leu Leu Lys Ile Ser Ser Ala Asn Lys
 325 330 335
 Tyr Phe Ser Arg Lys Tyr Ile His Met Ala Asn Ala Tyr Ile Ala Lys
 340 345 350
 Cys Ser Met Ser Ile Ile Thr Lys Lys Ala Ser Gln Lys Arg Lys Asn
 355 360 365
 Lys Thr Arg Arg Tyr Gln Leu Ile Pro Val Arg Met Thr Leu Ile Lys
 370 375 380
 Lys Lys Lys Arg Trp Ala Arg Cys Glu Glu Lys Gly Arg Leu Ala His
 385 390 395 400
 Cys Trp Phe Glu Cys Lys Ala Arg Gln Pro Leu Ala Lys Thr Lys Ala
 405 410 415
 Arg Phe Leu Lys Lys Leu Lys Leu Pro Cys His Thr Ala Ile Ala Leu
 420 425 430
 Leu Asp Ile Tyr Pro Lys Gln Ile Lys Ser Glu Ala Arg Asn Ile Cys
 435 440 445
 Asn Ser Val Tyr Ala Leu Phe Thr Ile Ala Lys Ile Gln Asn Lys Ser
 450 455 460
 Leu Thr Ser Asn Glu Ala Met Lys Thr Met Trp Ala Ile Tyr Thr Thr

77

465

470

475

480

Glu Tyr Tyr Phe Ala Asn Lys Lys Ile Pro Phe Leu
 485 490

<210> 213

<211> 37

<212> PRT

<213> Homo sapiens

<400> 213

Met Met Leu Pro Pro Asn Leu Glu Asn Thr Gly Ser His Ile Ser Pro
 1 5 10 15

Glu Trp Arg Phe Met Arg Arg Asn Thr Asn Glu Lys Lys Lys Trp Ser
 20 25 30

Met Lys Pro Glu Leu
 35

<210> 214

<211> 67

<212> PRT

<213> Homo sapiens

<400> 214

Met Cys His Glu Leu Trp Pro Cys Leu Tyr Phe Tyr Phe Asn Arg Asn
 1 5 10 15

His Leu Phe Lys Gln Lys Val Leu His Leu Asn Cys His Asn Cys Val
 20 25 30

Cys Val Ile Asn Ile Ser Tyr Phe Ile Gln Ala Gln Pro Thr Leu Ala
 35 40 45

Phe Ile Asn Ala His Asn Gln Glu Ile Asn Leu Ile Leu Thr Lys Asn
 50 55 60

Tyr Pro Ser
 65

<210> 215

<211> 12

<212> PRT

<213> Homo sapiens

<400> 215

Met Ser His Asn Ile Asp Leu Leu Gly Lys Asp Phe
 1 5 10

<210> 216

<211> 39

<212> PRT

<213> Homo sapiens

<400> 216

Met Arg Glu Cys Gly Glu Ser Ile Cys Pro Ser Leu Ala Gly His Arg

78

1 5 10 15
 Leu Ser Arg Gly Ala Val Glu Val Glu Thr Thr Gln Asp Ser Glu Ser
 20 25 30
 Pro Gln Val His Pro Gly Pro
 35

<210> 217

<211> 89

<212> PRT

<213> Homo sapiens

<400> 217

Met Leu Leu Ser Cys Cys Ser Gln Asn Gln Lys Met Ala Ser Arg Ser
 1 5 10 15
 Ala Gln Ser Ser Gln Glu Gln Met Leu Arg Val Thr Leu Glu Ser Phe
 20 25 30
 Cys Cys Leu His Ile Gln Thr Ile Thr Ile Ser Leu Ile Ser Leu Leu
 35 40 45
 Tyr Ile Phe His Met Cys Pro Leu Leu Ser Ile Cys Thr Leu Ile Ser
 50 55 60
 Glu Gly His Gln His Leu Ser Ser Glu Cys Leu Gln Tyr Leu Leu Thr
 65 70 75 80
 Gly His Gln Ala Ser Ser Phe Ala Pro
 85

<210> 218

<211> 56

<212> PRT

<213> Homo sapiens

<400> 218

Met Asp Cys Thr Ala Val Gly Arg Gly Thr Arg Arg Ala Ser Ala Pro
 1 5 10 15
 Thr Cys Glu Arg Arg Pro Arg Gly Leu Arg Cys Arg Arg Pro Val Ala
 20 25 30
 Pro Pro Pro Arg Ala Leu Ser Ala Val Asn Leu Gly Arg Arg Arg Trp
 35 40 45
 Gly Ser Gly Lys Arg Arg Ala Gln
 50 55

<210> 219

<211> 36

<212> PRT

<213> Homo sapiens

<400> 219

Ala Ala Ala Ala Pro Pro Pro Ala Pro Pro His His Gly Ala Ala Ala
 1 5 10 15

79

Pro Pro Pro Gly Gln Leu Ser Pro Ala Ser Pro Ala Thr Ala Ala Pro
 20 25 30

Pro Ala Pro Ala
 35

<210> 220

<211> 85

<212> PRT

<213> Homo sapiens

<400> 220

Met Ala Gly Pro Arg Cys Pro Arg Lys Gly Arg Thr Asn Thr Cys Val
 1 5 10 15

Cys Ser Ala Asn Pro Leu Glu Ala Val Gln Lys Pro Leu Ala Ala Gly
 20 25 30

Pro Thr Arg Arg Gly Gly Gly Trp Asp Pro Ala Gly Ala Gly Ala Ala
 35 40 45

Trp Leu His Gly Leu Tyr Ser Val Tyr Thr Ala Gly Gly Arg Gly Gly
 50 55 60

Arg Leu Arg Phe Leu Arg Tyr Gln Ser Arg Arg Phe Gly His Leu Arg
 65 70 75 80

Ala Pro Ala Ala Gly
 85

<210> 221

<211> 376

<212> PRT

<213> Homo sapiens

<400> 221

Met Met Ala Ser Tyr Pro Glu Pro Glu Asp Ala Ala Gly Ala Leu Leu
 1 5 10 15

Ala Pro Glu Thr Gly Arg Thr Val Lys Glu Pro Glu Gly Pro Pro Pro
 20 25 30

Ser Pro Gly Lys Gly Gly Gly Gly Gly Gly Thr Ala Pro Glu Lys
 35 40 45

Pro Asp Pro Ala Gln Lys Pro Pro Tyr Ser Tyr Val Ala Leu Ile Ala
 50 55 60

Met Ala Ile Arg Glu Ser Ala Glu Lys Arg Leu Thr Leu Ser Gly Ile
 65 70 75 80

Tyr Gln Tyr Ile Ile Ala Lys Phe Pro Phe Tyr Glu Lys Asn Lys Lys
 85 90 95

Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu Cys Phe
 100 105 110

Ile Lys Val Pro Arg Glu Gly Gly Gly Glu Arg Lys Gly Asn Tyr Trp

80

115	120	125
Thr Leu Asp Pro Ala Cys Glu Asp Met Phe Glu Lys Gly Asn Tyr Arg		
130	135	140
Arg Arg Arg Arg Met Lys Arg Pro Phe Arg Pro Pro Pro Ala His Phe		
145	150	155
Gln Pro Gly Lys Gly Leu Phe Gly Ala Gly Gly Ala Ala Gly Gly Cys		
	165	170
Gly Val Ala Gly Ala Gly Ala Asp Gly Tyr Gly Tyr Leu Ala Pro Pro		
	180	185
Lys Tyr Leu Gln Ser Gly Phe Leu Asn Asn Ser Trp Pro Leu Pro Gln		
	195	200
Pro Pro Ser Pro Met Pro Tyr Ala Ser Cys Gln Met Ala Ala Ala Ala		
	210	215
Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Pro Gly Ser Pro Gly		
225	230	235
Ala Ala Ala Val Val Lys Gly Leu Ala Gly Pro Ala Ala Ser Tyr Gly		
	245	250
Pro Tyr Thr Arg Val Gln Ser Met Ala Leu Pro Pro Gly Val Val Asn		
	260	265
Ser Tyr Asn Gly Leu Gly Gly Pro Pro Ala Ala Pro Pro Pro Pro Pro		
	275	280
His Pro His Pro His Pro His Ala His His Leu His Ala Ala Ala Ala		
	290	295
Pro Pro Pro Ala Pro Pro His His Gly Ala Ala Ala Pro Pro Pro Gly		
305	310	315
Gln Leu Ser Pro Ala Ser Pro Ala Thr Ala Ala Pro Pro Ala Pro Ala		
	325	330
Pro Thr Ser Ala Pro Gly Leu Gln Phe Ala Cys Ala Arg Gln Pro Glu		
	340	345
Leu Ala Met Met His Cys Ser Tyr Trp Asp His Asp Ser Lys Thr Gly		
	355	360
Ala Leu His Ser Arg Leu Asp Leu		
	370	375

<210> 222

<211> 19

<212> PRT

<213> Homo sapiens

<400> 222

Met Gln Tyr Phe Ser Leu Pro Val Leu Thr Leu Leu Met Val Pro Phe
1 5 10 15

Ile Phe Ile

<210> 223
<211> 30
<212> PRT
<213> Homo sapiens

<400> 223
Met Pro Leu Lys His Ile Lys Phe Lys Asn Leu Phe Leu Leu Ala Leu
1 5 10 15
Glu Ile Leu Trp Asn Phe Thr Trp Asn Leu Ile Leu Gly Arg
20 25 30

<210> 224
<211> 52
<212> PRT
<213> Homo sapiens

<400> 224
Met Leu Ile Met Lys Glu Thr His Glu Gln Leu Ser Glu Glu Ser Gly
1 5 10 15
Glu Val Gly Met Ile Ser Glu His Arg Gly Gly Ser Pro Ala Trp Gly
20 25 30
Leu Pro Asn Pro Asp Ala Gln Lys Phe Leu Ser Arg Pro His Tyr Thr
35 40 45
Gly Met Ile Asp
50

<210> 225
<211> 52
<212> PRT
<213> Homo sapiens

<400> 225
Met Gly Leu Asn Pro Gly Val Cys Leu Glu Pro Gln Leu Val Cys Asp
1 5 10 15
Thr Asp His His Phe Leu Lys Thr Ile Tyr Lys Asn Lys Thr Arg Cys
20 25 30
Met Lys Phe Arg Phe Trp Lys Lys Val Gln Val Phe Met Asn Ile Ser
35 40 45
Glu Leu Pro Lys
50

<210> 226
<211> 19
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE

82

<222> (14)

<220>

<221> UNSURE

<222> (18)

<400> 226

Met Asp Asn Glu Asn Gln Asn Ile Lys Lys Glu Lys Lys Xaa Lys Lys
 1 5 10 15

Lys Xaa Lys

<210> 227

<211> 75

<212> PRT

<213> Homo sapiens

<400> 227

Phe Phe Phe Leu Arg Gln Ser Leu Ala Leu Ser Pro Arg Leu Glu Cys
 1 5 10 15

Ser Gly Ala Ile Ser Ala His Cys Lys Leu Arg Leu Pro Gly Ser Cys
 20 25 30

His Phe Pro Ala Ser Ala Ser Gln Val Ala Glu Thr Thr Gly Thr Arg
 35 40 45

His Asn Ala Arg Val Ile Phe Cys Ile Leu Val Glu Thr Gly Phe His
 50 55 60

Arg Val Ser Gln Asp Gly Leu Asp Leu Leu Thr
 65 70 75

<210> 228

<211> 95

<212> PRT

<213> Homo sapiens

<400> 228

Met Arg Arg Ala Lys Ala Pro Lys Ile Arg Gly Thr Ala Asn Ala Thr
 1 5 10 15

Asp Arg Lys Lys Ala Glu Gly Lys Ser Ala Ser Ser Arg Leu Arg Pro
 20 25 30

Arg Gly Pro Ala Leu Ala Pro Ala Ser Ile His Arg Glu His Thr Gln
 35 40 45

Glu Ala Phe Glu Trp Pro Gly Phe Leu Val Ser Leu Ala Gln Arg Gln
 50 55 60

Glu Leu Glu His Glu Arg Ser Ser Glu Thr Leu Trp Val Leu Pro Thr
 65 70 75 80

Leu Arg Gln Ala Ser Gln His Leu His Ala Leu Leu Cys Ser Pro
 85 90 95

83

<210> 229
 <211> 98
 <212> PRT
 <213> Homo sapiens

<400> 229
 Met Val Gly Ala Ser Pro Gly Gly Met Gly Cys Glu Gly Gly Arg Met
 1 5 10 15
 Arg Ala Arg Arg Phe Ser Leu Gly Asp Pro Ala Thr Gln Ser His Leu
 20 25 30
 Pro Leu Thr Glu Gly Ser Arg Ala Pro Ser Gly Pro Leu Ala Thr Lys
 35 40 45
 Ala Gln Leu Lys Ser Gln Lys Gly His Ile Arg Ser Gln Ala Thr Gly
 50 55 60
 Thr Ala His Val Arg Asn Val Ser Ala Met Glu Lys Tyr Lys Thr Arg
 65 70 75 80
 Lys Glu Val Cys Gly Pro Asn Arg Thr Cys Leu Ser Thr Phe Tyr Cys
 85 90 95
 Asn Val

<210> 230
 <211> 84
 <212> PRT
 <213> Homo sapiens

<400> 230
 Met Asp Thr Thr Asn Asn Gln Ile Asn Leu Tyr Ile His Thr Lys Phe
 1 5 10 15
 Phe Leu Lys Ile Lys Val Asn Thr Ser Ile Ser Lys Arg Leu Phe Ser
 20 25 30
 Pro Tyr Phe Asn Ile His Ile Phe Cys Met Phe Ile Tyr Val His Gly
 35 40 45
 Gly Cys Phe Tyr Ile Pro Arg Lys Phe Arg Cys Tyr Ser Arg Arg Leu
 50 55 60
 Ser Ile Ile His Thr Ala Val Lys Trp Ser Pro Ala Leu Ser Arg His
 65 70 75 80
 Pro Thr Ala Gln

<210> 231
 <211> 924
 <212> PRT
 <213> Homo sapiens

<400> 231
 Gly Arg Leu Thr Phe Arg Asp Val Ala Ile Glu Phe Ser Leu Ala Glu
 1 5 10 15

Trp Lys Cys Leu Asn Pro Ser Gln Arg Ala Leu Tyr Arg Glu Val Met
 20 25 30
 Leu Glu Asn Tyr Arg Asn Leu Glu Ala Val Asp Ile Ser Ser Lys Arg
 35 40 45
 His Asp Glu Gly Gly Leu Val Asn Arg Ala Arg Gln Tyr Arg Ser Asp
 50 55 60
 Pro His Arg Asp Ile Ala Lys Ile Ser Lys Leu Ser His Trp Arg Phe
 65 70 75 80
 Leu Leu Pro Gly Asn Ala Glu Arg Asn Ser Ala Tyr Ala Val Ser Val
 85 90 95
 Ser Arg Arg Glu Arg Asn Gly His Glu Ala Pro Met Thr Lys Ile Lys
 100 105 110
 Lys Leu Thr Gly Ser Thr Asp Gln His Asp His Arg His Ala Gly Asn
 115 120 125
 Lys Pro Ile Lys Asp Gln Leu Gly Ser Ser Phe Tyr Ser His Leu Pro
 130 135 140
 Glu Leu His Ile Ile Gln Ile Lys Gly Lys Ile Gly Asn Gln Phe Glu
 145 150 155 160
 Lys Ser Thr Ser Asp Ala Pro Ser Val Ser Thr Ser Gln Arg Ile Ser
 165 170 175
 Pro Arg Pro Gln Ile His Ile Ser Asn Asn Tyr Gly Asn Asn Ser Pro
 180 185 190
 Asn Ser Ser Leu Leu Pro Gln Lys Gln Glu Val Tyr Met Arg Glu Lys
 195 200 205
 Ser Phe Gln Cys Asn Glu Ser Gly Lys Ala Phe Asn Cys Ser Ser Leu
 210 215 220
 Leu Arg Lys His Gln Ile Pro His Leu Gly Asp Lys Gln Tyr Lys Cys
 225 230 235 240
 Asp Val Cys Gly Lys Leu Phe Asn His Lys Gln Tyr Leu Thr Cys His
 245 250 255
 Arg Arg Cys His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys Gly
 260 265 270
 Lys Ser Phe Ser Gln Val Ser Ser Leu Thr Cys His Arg Arg Leu His
 275 280 285
 Thr Ala Val Lys Ser His Lys Cys Asn Glu Cys Gly Lys Ile Phe Gly
 290 295 300
 Gln Asn Ser Ala Leu Val Ile His Lys Ala Ile His Thr Gly Glu Lys
 305 310 315 320
 Pro Tyr Lys Cys Asn Glu Cys Asp Lys Ala Phe Asn Gln Gln Ser Asn
 325 330 335

85

Leu Ala Arg His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys
 340 345 350
 Glu Glu Cys Asp Lys Val Phe Ser Arg Lys Ser Thr Leu Glu Ser His
 355 360 365
 Lys Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Lys Val Cys Asp
 370 375 380
 Thr Ala Phe Thr Trp Asn Ser Gln Leu Ala Arg His Lys Arg Ile His
 385 390 395 400
 Thr Gly Glu Lys Thr Tyr Lys Cys Asn Glu Cys Gly Lys Thr Phe Ser
 405 410 415
 His Lys Ser Ser Leu Val Cys His His Arg Leu His Gly Gly Glu Lys
 420 425 430
 Ser Tyr Lys Cys Lys Val Cys Asp Lys Ala Phe Ala Trp Asn Ser His
 435 440 445
 Leu Val Arg His Thr Arg Ile His Ser Gly Gly Lys Pro Tyr Lys Cys
 450 455 460
 Asn Glu Cys Gly Lys Thr Phe Gly Gln Asn Ser Asp Leu Leu Ile His
 465 470 475 480
 Lys Ser Ile His Thr Gly Glu Gln Pro Tyr Lys Tyr Glu Glu Cys Glu
 485 490 495
 Lys Val Phe Ser Cys Gly Ser Thr Leu Glu Thr His Lys Ile Ile His
 500 505 510
 Thr Gly Glu Lys Pro Tyr Lys Cys Lys Val Cys Asp Lys Ala Phe Ala
 515 520 525
 Cys His Ser Tyr Leu Ala Lys His Thr Arg Ile His Ser Gly Glu Lys
 530 535 540
 Pro Tyr Lys Cys Asn Glu Cys Ser Lys Thr Phe Arg Leu Arg Ser Tyr
 545 550 555 560
 Leu Ala Ser His Arg Arg Val His Ser Gly Glu Lys Pro Tyr Lys Cys
 565 570 575
 Asn Glu Cys Ser Lys Thr Phe Ser Gln Arg Ser Tyr Leu His Cys His
 580 585 590
 Arg Arg Leu His Ser Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys Gly
 595 600 605
 Lys Thr Phe Ser His Lys Pro Ser Leu Val His His Arg Arg Leu His
 610 615 620
 Thr Gly Glu Lys Ser Tyr Lys Cys Thr Val Cys Asp Lys Ala Phe Val
 625 630 635 640
 Arg Asn Ser Tyr Leu Ala Arg His Thr Arg Ile His Thr Ala Glu Lys
 645 650 655
 Pro Tyr Lys Cys Asn Glu Cys Gly Lys Ala Phe Asn Gln Gln Ser Gln

86

660										665										670											
Leu	Ser	Leu	His	His	Arg	Ile	His	Ala	Gly	Glu	Lys	Leu	Tyr	Lys	Cys	Leu	Ser	Leu	His	His	Arg	Ile	His	Ala	Gly	Glu	Lys	Leu	Tyr	Lys	Cys
		675					680					685																			
Glu	Thr	Cys	Asp	Lys	Val	Phe	Ser	Arg	Lys	Ser	His	Leu	Lys	Arg	His	Glu	Thr	Cys	Asp	Lys	Val	Phe	Ser	Arg	Lys	Ser	His	Leu	Lys	Arg	His
	690					695					700																				
Arg	Arg	Ile	His	Pro	Gly	Lys	Lys	Pro	Tyr	Lys	Cys	Lys	Val	Cys	Asp	Arg	Arg	Ile	His	Pro	Gly	Lys	Lys	Pro	Tyr	Lys	Cys	Lys	Val	Cys	Asp
705					710					715					720	705					710					715					720
Lys	Thr	Phe	Gly	Ser	Asp	Ser	His	Leu	Lys	Gln	His	Thr	Gly	Leu	His	Lys	Thr	Phe	Gly	Ser	Asp	Ser	His	Leu	Lys	Gln	His	Thr	Gly	Leu	His
				725					730					735						725					730					735	
Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Asn	Glu	Cys	Gly	Lys	Ala	Phe	Ser	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Asn	Glu	Cys	Gly	Lys	Ala	Phe	Ser
		740						745					750							740					745					750	
Lys	Gln	Ser	Thr	Leu	Ile	His	His	Gln	Ala	Val	His	Gly	Val	Gly	Lys	Lys	Gln	Ser	Thr	Leu	Ile	His	His	Gln	Ala	Val	His	Gly	Val	Gly	Lys
	755					760						765																			
Leu	Asp	Ala	Cys	Asn	Asp	Cys	His	Lys	Val	Phe	Ser	Asn	Ala	Thr	Thr	Leu	Asp	Ala	Cys	Asn	Asp	Cys	His	Lys	Val	Phe	Ser	Asn	Ala	Thr	Thr
	770					775					780																				
Ile	Ala	Asn	His	Trp	Arg	Ile	Tyr	Asn	Glu	Ala	Arg	Ser	Asn	Lys	Cys	Ile	Ala	Asn	His	Trp	Arg	Ile	Tyr	Asn	Glu	Ala	Arg	Ser	Asn	Lys	Cys
785					790					795					800	785					790					795					800
Asn	Lys	Cys	Gly	Lys	Phe	Phe	Arg	His	His	Ser	Tyr	Ile	Ala	Val	His	Asn	Lys	Cys	Gly	Lys	Phe	Phe	Arg	His	His	Ser	Tyr	Ile	Ala	Val	His
				805					810					815						805					810					815	
Ala	His	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	His	Asp	Cys	Gly	Ala	His	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	His	Asp	Cys	Gly
			820					825					830							820					825					830	
Lys	Val	Phe	Ser	Gln	Ala	Ser	Ser	Tyr	Ala	Lys	His	Arg	Arg	Ile	His	Lys	Val	Phe	Ser	Gln	Ala	Ser	Ser	Tyr	Ala	Lys	His	Arg	Arg	Ile	His
	835						840					845								835											
Thr	Gly	Glu	Lys	Pro	His																										

```
<210> 232
<211> 322
<212> PRT
<213> Homo sapiens
```

<220>
<221> UNSURE
<222> (291) .. (299)

<400> 232

87

Met Leu Ala Ala Cys Leu Met Thr Pro Asp His Pro Thr Ala Gly Asn
 1 5 10 15
 Gln Pro Leu Arg Thr Pro Ser His Val Pro Gly Thr Cys Arg Cys Arg
 20 25 30
 Ser Gln His Pro Ala Val Trp Ala Leu Tyr Asp Asp Gln Leu Gly Asn
 35 40 45
 Val Gly Asp His His Val Ala Thr His Met Val Gly Pro His Asp His
 50 55 60
 Ile Leu Pro Ile Leu Gln Leu Leu Leu Pro Gly Asp Leu Arg Pro Gly
 65 70 75 80
 Pro Ala His His Ile Thr Glu Glu Thr His Cys Leu Thr His Gly Asp
 85 90 95
 Arg Leu Val His Thr Val Val Glu Gln Arg Arg Asp Arg His Val Gln
 100 105 110
 Leu Arg Gly Leu Trp Gly Gly Cys Ala Gly Val His Gly Gly Leu Arg
 115 120 125
 Cys Trp Gly Ala Gly Val Gly Pro Gly Glu Val Ile Ala Ala Gly Tyr
 130 135 140
 Asn Gly Gln Cys Asp Ala Phe Gly Ala Gly Leu Gly Ile His Val Ala
 145 150 155 160
 Ala Val Ile Val Gly Glu Ala Val Arg Gly Ala Gly Lys Ala Gly Leu
 165 170 175
 Leu Leu Thr Ala Val Phe Ala Leu Thr His Gly Leu Ala Ile Pro Asp
 180 185 190
 Val Thr Leu Arg Ala Leu Leu Gln Thr His Glu Val Val Thr Cys Gly
 195 200 205
 Leu Leu Gly His Ala His Trp Ala Leu Leu Pro Phe His Val His Val
 210 215 220
 Ala Gly Arg His Ala Ala Leu Gly Pro Thr Tyr Val Gly Ala Ala Leu
 225 230 235 240
 Leu Ile Gly Leu Thr Leu Leu Val Arg Leu Thr Leu Pro Pro Ala Gly
 245 250 255
 Ala Leu Cys Val His Pro Glu Val Gly Ile His Val Val Gly Ala Asp
 260 265 270
 Ala Gly Val Gly Ile Ala Asp Gly Arg Gln Arg Gln Ala Ser Arg Gly
 275 280 285
 His Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys His Leu Leu Pro
 290 295 300
 Ala Arg Pro Glu Pro Ala Thr Pro Trp Gly Pro His Gly Ala Gly Trp
 305 310 315 320
 Gly Gly

<210> 233

<211> 503

<212> PRT

<213> Homo sapiens

<400> 233

Glu Cys Glu Thr Tyr Glu Lys Cys Cys Pro Asn Val Cys Gly Thr Lys
 1 5 10 15

Ser Cys Val Ala Ala Arg Tyr Met Asp Val Lys Gly Lys Lys Gly Pro
 20 25 30

Val Gly Met Pro Lys Glu Ala Thr Cys Asp His Phe Met Cys Leu Gln
 35 40 45

Gln Gly Ser Glu Cys Asp Ile Trp Asp Gly Gln Pro Val Cys Lys Cys
 50 55 60

Lys Asp Arg Cys Glu Lys Glu Pro Ser Phe Thr Cys Ala Ser Asp Gly
 65 70 75 80

Leu Thr Tyr Tyr Asn Arg Cys Tyr Met Asp Ala Glu Ala Cys Ser Lys
 85 90 95

Gly Ile Thr Leu Ala Val Val Thr Cys Arg Tyr His Phe Thr Trp Pro
 100 105 110

Asn Thr Ser Pro Pro Ala Pro Glu Thr Thr Met His Pro Ser Thr Ala
 115 120 125

Ser Pro Glu Thr Pro Glu Leu Asp Met Ala Val Pro Ala Leu Leu Asn
 130 135 140

Asn Arg Val His Gln Ser Val Thr Met Gly Glu Thr Val Ser Phe Leu
 145 150 155 160

Cys Asp Val Val Gly Arg Pro Arg Pro Glu Ile Thr Trp Glu Lys Gln
 165 170 175

Leu Glu Asp Arg Glu Asn Val Val Met Arg Pro Asn His Val Arg Gly
 180 185 190

Asn Val Val Val Thr Asn Ile Ala Gln Leu Val Ile Tyr Asn Ala Arg
 195 200 205

Leu Gln Asp Ala Gly Ile Tyr Thr Cys Thr Ala Arg Asn Val Ala Gly
 210 215 220

Val Leu Arg Ala Asp Phe Pro Leu Ser Asp Gly Gln Gly Ser Ser Gly
 225 230 235 240

Met Gln Pro Ala Ser Glu Ser Ser Pro Asn Gly Thr Ala Phe Pro Ala
 245 250 255

Ala Glu Cys Leu Lys Pro Pro Asp Ser Glu Asp Cys Gly Glu Glu Gln
 260 265 270

Thr Arg Trp His Phe Asp Ala Gln Ala Asn Asn Cys Leu Thr Phe Thr

89

275					280					285					
Phe	Gly	His	Cys	His	Arg	Asn	Leu	Asn	His	Phe	Glu	Thr	Tyr	Glu	Ala
290					295					300					
Cys	Met	Leu	Ala	Cys	Met	Ser	Gly	Pro	Leu	Ala	Ala	Cys	Ser	Leu	Pro
305					310					315					320
Ala	Leu	Gln	Gly	Pro	Cys	Lys	Ala	Tyr	Ala	Pro	Arg	Trp	Ala	Tyr	Asn
				325					330					335	
Ser	Gln	Thr	Gly	Gln	Cys	Gln	Ser	Phe	Val	Tyr	Gly	Gly	Cys	Glu	Gly
			340					345					350		
Asn	Gly	Asn	Asn	Phe	Glu	Ser	Arg	Glu	Ala	Cys	Glu	Glu	Ser	Cys	Pro
		355					360					365			
Phe	Pro	Arg	Gly	Asn	Gln	Arg	Cys	Arg	Ala	Cys	Lys	Pro	Arg	Gln	Lys
	370					375					380				
Leu	Val	Thr	Ser	Phe	Cys	Arg	Ser	Asp	Phe	Val	Ile	Leu	Gly	Arg	Val
385					390					395					400
Ser	Glu	Leu	Thr	Glu	Glu	Pro	Asp	Ser	Gly	Arg	Ala	Leu	Val	Thr	Val
				405					410					415	
Asp	Glu	Val	Leu	Lys	Asp	Glu	Lys	Met	Gly	Leu	Lys	Phe	Leu	Gly	Gln
			420					425					430		
Glu	Pro	Leu	Glu	Val	Thr	Leu	Leu	His	Val	Asp	Trp	Ala	Cys	Pro	Cys
		435					440					445			
Pro	Asn	Val	Thr	Val	Ser	Glu	Met	Pro	Leu	Ile	Ile	Met	Gly	Glu	Val
		450					455					460			
Asp	Gly	Gly	Met	Ala	Met	Leu	Arg	Pro	Asp	Ser	Phe	Val	Gly	Ala	Ser
465						470					475				480
Ser	Ala	Arg	Arg	Val	Arg	Lys	Leu	Arg	Glu	Val	Met	His	Lys	Lys	Thr
				485					490					495	
Cys	Asp	Val	Leu	Lys	Glu	Phe									
			500												

<210> 234

<211> 89

<212> PRT

<213> Homo sapiens

<400> 234

Met	Phe	Leu	Phe	Leu	Leu	Gln	Pro	Pro	Pro	Ser	Ser	Leu	Ser	Pro	Leu
1				5					10					15	
Leu	Pro	Pro	Ser	Leu	Pro	Ala	Phe	Ser	Ser	Ser	Phe	Ile	Ser	Pro	Ala
			20					25					30		
Thr	Lys	Gln	Ile	Pro	Gly	Leu	Leu	Ser	Asp	Leu	Cys	Pro	Arg	Lys	Pro
		35					40					45			
Val	Ala	Tyr	Glu	Ser	Thr	Pro	Ser	Ile	Arg	Gln	Lys	Leu	Gln	Thr	Val

90

50

55

60

Val Ser Pro Ala Glu Gly Cys Val Trp Gly Pro Trp Asp Glu Gly Ile
 65 70 75 80

Cys Val Gly Ala Leu Arg Thr Gly Gln
 85

<210> 235

<211> 29

<212> PRT

<213> Homo sapiens

<400> 235

Met Gly Gly Ala Leu Leu Pro Pro Asp Arg Asp Glu Ser Pro Arg Tyr
 1 5 10 15

Leu Leu Asn Leu Cys Asn Thr Pro Ala Gly Lys Leu Gly
 20 25

<210> 236

<211> 38

<212> PRT

<213> Homo sapiens

<400> 236

Met Pro Ser Leu Ser Glu Ser Ile Leu Leu Ser Ser Glu Val Cys Asp
 1 5 10 15

Trp Thr Lys Leu Ser Thr Ile Phe Ser Ser Ala Asn Asn Leu Leu Leu
 20 25 30

Ile Cys Cys Lys Val Ser
 35

<210> 237

<211> 33

<212> PRT

<213> Homo sapiens

<400> 237

Met Leu Pro Ser Gly Val Lys Lys Phe Phe Val Asp Arg Ala Phe Glu
 1 5 10 15

Leu Arg Ser Phe Lys Tyr Thr Thr Asp Val Pro Leu Arg Glu Thr Asp
 20 25 30

Leu

<210> 238

<211> 88

<212> PRT

<213> Homo sapiens

<400> 238

Met Gln Ala Ser Pro Leu Gln Ile Arg Gln Asn Pro Ala Leu Phe Leu

91

1				5					10					15	
Val	Met	Thr	Phe	Pro	Thr	Ala	Arg	Gly	His	Lys	Ser	Met	Ile	Gln	His
			20					25					30		
Tyr	Arg	Asn	Pro	Pro	Thr	Ser	Arg	Lys	Val	Ser	Thr	Thr	His	Lys	Asp
		35					40					45			
Ser	His	Val	His	Ala	Asp	Thr	Lys	Thr	His	Phe	Arg	Glu	Glu	Ala	Pro
	50					55					60				
Arg	His	Ser	Leu	Lys	Pro	Gln	Leu	Gly	Thr	Phe	Leu	His	Asp	Asn	Ser
65					70					75				80	
Ser	Ala	Ser	Leu	Gly	Gln	Cys	Asn								
				85											

1

94